MODELING RELIABILITY IN COPPER/LOW-K INTERCONNECTS AND VARIABILITY IN CMOS

A Dissertation
Presented to
The Academic Faculty

by

Muhammad Muqarrab Bashir

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in
Electrical and Computer Engineering

School of Electrical and Computer Engineering
Georgia Institute of Technology
August 2011

MODELING RELIABILITY IN COPPER/LOW-K INTERCONNECTS AND VARIABILITY IN CMOS

Approved by:

Dr. Linda Milor, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Sung K. Lim
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Abhijit Chatterjee
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. John Barry
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Kobi Abayomi
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: May 13, 2011

To my parents Dr. Muhammad Bashir Sadiq and Musarrat.

ACKNOWLEDGEMENTS

I would like to thank my advisor Linda Milor for her guidance and her efforts towards my development as a researcher. I would also like to thank Kristof Croes from IMEC, Belgium, Gautam Verma from Altera Corporation, and J.R. Lloyd from SUNY, Albany, for their valuable input. I would also like to thank Kim Dae Hyun, for developing the layout extractor for the reliability simulator, and Krit Athikulwonsge, for temperature maps, from the GTCAD lab.

Muhammad Muqarrab Bashir

Atlanta,

May, 2011

TABLE OF CONTENTS

PART II VARIABILITY

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC	Alternating Current
Al	Aluminum
Ar	Argon
ARDE	Aspect Ratio Dependent Etching
BEOL	Backend of the Line
BTI	Bias Temperature Instability
BTS	Bias Temperature Stress
C	Carbon
C-V	Capacitance Voltage
CD	Critical Dimension
CMOS	Complimentary Metal Oxide Semiconductor
CMP	Chemical-Mechanical Polishing
Cu	Copper
CVS	Constant Voltage Stress
DC	Direct Current
DFM	Design for Manufacturability
DIBL	Drain-Induced Barrier Lowering
E	Electric Field

EM	Electromigration
F	Fluorine
FEM	Finite Element Modeling
HF	High Fields
I-V	Current Voltage
IC	Integrated Circuit
ILD	Intra-Level Dielectrics
ITRS	International Technology Roadmap for Semiconductors
LER	Line Edge Roughness
MOS	Metal Oxide Silicon
MOSFET	Metal-Oxide-Semiconductor-Field Effect Transistor
MTTF	Mean-Time-to-Failure
NMOS	N-Channel MOSFET
NTRS	National Technology Roadmap for Semiconductors
OAI	Off-Axis Illumination
OPC	Optical Proximity Correction
PCA	Principal Component Analysis
PDF	Probability Density Function
PF	Poole-Frenkel Emission
PSM	Phase-Shift Masks

RC	Resistive-Capacitive
S	Linespace
SE	Schottky Emission
SEM	Scanning Electron Microscope
Si	Silicon
SILC	Stress Induced Leakage Current
SiO_2	Silicon Dioxide
SRAM	Static Random Access Memories
STEM	Scanning Transmission Electron Microscopy
T	Temperature
Ta	Tantalum
TaN	Tantalum Nitride
TDDB	Time-Dependent Dielectric Breakdown
TEM	Transmission Electron Microscope
TEOS	Tetraethyl Orthosilicate
TF	Time-to-Failure
V	Voltage
VR	Voltage Ramp

SUMMARY

The objective of the research is to model the impact of physical design characteristics on backend dielectric reliability. Backend dielectric breakdown is a concern for a number of reasons. The impact of process on backend dielectric reliability is significant, and well modeled, however; the impact of physical design and geometry on backend low-k must also be modeled since the failure mechanism is field driven, and thus impacted by the interaction of operation and physical design. This research identifies physical design parameters that are crucial to backend dielectric breakdown. The impact of these physical design parameters on backend low-k time-dependent dielectric breakdown will be modeled. The insights gathered from this work will then be unified into a methodology to predict chip backend low-k time-dependent dielectric breakdown reliability. Thus, a methodology to predict chip reliability will be proposed.

Variation in the manufacturing process results in variability in device and circuit characteristics. Understanding and modeling within-die variation based on process data can aid in circuit design by enabling the development of techniques to compensate for such variation and to properly design in the presence of within-die variation. This research proposes a methodology to model variation in device parameters and characteristics. New methods of electrical and physical parameter extraction will be proposed. Extracted values will be used to develop models of variation in physical and electrical parameters of transistors.

CHAPTER I

INTRODUCTION

1.1 Background

1.1.1 Backend Low-k Time-Dependent Dielectric Breakdown Reliability

With the shrinking of integrated circuit (IC) dimensions, the density of active devices on chip increased, resulting in a decrease in the switching speed of the metal-oxide-silicon (MOS) devices. This decrease was accompanied by an increase in the resistive-capacitive (RC) delay introduced by the interconnect lines. This problem was further exacerbated by the decrease in device delay with each successive generation and the concomitant increase in the RC delay of interconnects [1]. RC delay increases approximately with the square of the scaling factor [2, 3]. RC delay is determined by the material properties of the metal, the interconnect dielectric, and the dimensions of the metal wire. Specifically [4]

$$\tau = 2\rho k\epsilon_0(\frac{4L^2}{P^2} + \frac{4L^2}{T^2}), \tag{1.1}$$

where τ is the RC delay and k is the dielectric constant of the dielectric material used, ρ is the resistivity of the metal, ϵ_0 is the permittivity of free space, L is the length of interconnect, P is the pitch of interconnect, and T is the line thickness. Reduction in T and P will result in an increase in τ. Line lengths will only increase as integration densities increase. Thus the only feasible solution to reduce the RC delay was to use metals with resistivity lower than that of Aluminum (Al) and dielectrics with lower k than Silicon Dioxide (SiO_2). Copper (Cu) provided an attractive alternative to Al because of its lower resistivity, better resistance against electromigration (EM), and potential for higher current densities. Moreover, dielectrics with lower k values meant a reduction in coupling capacitance between neighboring interconnect lines

and improved power dissipation properties [5]. Even with an SiO_2 dielectric, the use of Cu interconnect is effectively equivalent to changing the SiO_2 dielectric, k of 3.9, with a dielectric having a k of 3 [6].

1.1.1.1 Cu Low-k Interconnects

When Cu interconnects are used with a low-k dielectric as intra-level dielectric (ILD), the solution packaged together has become to known as Cu/low-k interconnects.

SILK [7,8] and XeroGel [9] were used as low-k dielectrics in the earliest Cu/low-k interconnect systems. However, the use of Cu metallization with low-k ILDs gave rise to new integration challenges. Despite the integration challenges, the National Technology Roadmap for Semiconductors (NTRS) specified the implementation of low-k ILD having dielectric constant of 2–2.5 by 2001 [10]. However, this target was revised in 1999 to use materials with k of 2.7–3.5 by 2001 for 0.18 micron technology [11]. The delay in the inevitable implementation of Cu/low-k interconnects can be attributed to integration issues [12].

1.1.1.2 Cu/low-k Interconnects: Integration and Reliability

Aluminum metallization dominated the Silicon (Si) based solid-state components because of its material properties, low resistivity, and the ease of integration. Al showed compatibility with Si and SiO_2 [13,14] because of the formation of a thin layer of Aluminium Oxide (Al_2O_3) at the interface of the two materials, Al (or Al alloys) and SiO_2. Cu/low-k interconnects, on the other hand, complicate process integration. Cu metallization requires the use of a barrier layer to prevent Cu diffusion into the ILD. Cu, susceptible to chemical-mechanical polishing (CMP), needs etch-stop and capping layers post-CMP. Moreover, Cu shows poor adhesion properties with SiO_2.

Low-k dielectrics show poor resistance against mechanical and thermal fatigue. Cu acts as a charged impurity in the dielectric film. Metal atoms, originating from the metal electrode, are known to diffuse interstitially, showing solubility in thin films

under high electric fields [15].

Polymer based low-k materials show charge instability when placed in direct contact with either metal or Si. These charge instabilities exacerbate over time under electric fields. The occurrence of these instabilities at room temperature indicates the involvement of electrons and holes. Degradation in capacitance voltage (C-V) characteristics of Cu interconnects is observed after bias-temperature-stress (BTS) whereas the same wasn't observable in Al interconnects [16]. Cu dissolves at a slower rate in Benzocycolbutane-based low-k polymers. However, Benzocycolbutane-based polymers reduced the process leverage because of a lower process temperature ceiling [17]. This in turn requires effective sealing of Cu in nitrides or oxides [4]. Tantalum (Ta) and Tantalum Nitride(TaN) could both effectively work as a barrier against Cu diffusion, but TaN forms a superior barrier layer than Ta [18]. Moreover, the amount of Cu left behind the barrier should be well regulated to control the breakdown performance [19]. Organic aromatic low-k materials show acceptable thermal stability and good dielectric barrier properties. However, they require a much thicker adhesion promotion layer, exhibit poor electrical characteristics, and show unstable C-V characteristics [20]. Not only are the barrier layers important to the operation of Cu/low-k dielectrics, but etch-stop layers have also been found to be critical because of their impact on leakage components [21]. Even the presence of a barrier cannot stop the diffusion of Cu ion in the dielectric [19, 22]. Breakdown characteristics are strongly dependent on the condition of the Cu surface and Tetraethyl-Orthosilicate (TEOS) capping layer before being capped by the Silicon Nitride (SiN) layer [23]. This may be one of the reasons for the presence of vulnerable sites near the CMP surfaces [24].

Low-k materials have poor thermal characteristics, which contribute to an increased thermal impedance [25]. Joule heating considerations require dielectrics with sufficiently large thermal conductivity [26]. Moreover, low-k materials show smaller

values of Young's modulus and tensile strength and consequently poor mechanical and stress characteristics. Compliance of dielectrics above their glass transition temperature has been shown to cause buckling and delamination of the capping film [27].

1.1.1.3 Time-Dependent Dielectric Breakdown

Dielectric breakdown is the irreversible local breakdown of a dielectric's insulation property. Time-dependent dielectric breakdown (TDDB) is dielectric breakdown that takes place after a constant application of an electric field (E), lower than the breakdown field, to the dielectric. TDDB results in the local development of very small spot with increased conductivity compared to the rest of the dielectric resulting in a change in the electrical characteristics of the dielectric [28].

In device operation, TDDB of the gate oxide renders the device useless because of the change in current properties. In interconnects, TDDB of the low-k dielectric leads to a catastrophic breakdown of the system.

1.1.1.4 Physical Mechanisms of Time-Dependant Dielectric Breakdown

The investigation of integrated circuit dielectric breakdown was initially motivated by complimentary metal oxide semiconductor (CMOS) gate dielectric breakdown and focused on SiO_2. Earlier work by researchers indicated that mobile ions caused breakdown in thin films. Accumulation of mobile ions at the metal-dielectric interface produced a barrier lowering, leading to eventual time-dependant dielectric breakdown [29, 30]. Other studies explained the breakdown as an avalanche process triggered by an electron with sufficient energy [31]. Some studies also argued that the breakdown occurred through impact ionization [32, 33]. High electric fields, along with conditions favoring electron injection in the film, gave rise to high defect densities near the injecting electrode [34]. The failure distributions for these breakdowns were described as extrinsic, because the cause of the failure were extrinsic defects.

Advancement in CMOS processing technology led to the elimination of mobile

ions from the standard NMOS process flow. Despite the absence of mobile ions, the phenomenon of breakdown was still observable. As a result, research focusing on the reliability of thin films shifted toward seeking intrinsic causes of breakdown. The statistical nature of the breakdown was identified [35, 36], along with the need for accurate statistical characterization of TDDB [35].

TDDB failure times were shown to relate directly to electric field, operating temperature (T) [37, 38], and voltage (V) [39]. However, it is not possible, *a priori*, to determine whether or not a given operating voltage will raise dielectric reliability issues or not [40].

The general pattern followed by studies aimed at breakdown distributions of a population of test structures has been to use either a constant or ramp voltage under constant temperature conditions or at different temperatures. The data obtained from these studies is then modeled using either the Weibull distribution or the Log-Normal distribution to find characteristics of the life time data [41]. The Log-Normal distribution is used to model failure mechanisms that are not restricted to a small localized spot and the damage is fairly extensive in nature, having started from a small localized spot that grew until failure. The Weibull distribution is a form of extreme value distribution that describes breakdown resulting from the weakest link of the system, i.e., the constituent of the system that fails earliest and leads to the failure of the system. The Weibull distribution can describe, through its shape parameter, the complete reliability bath tub curve. Different values of the shape parameter of the Weibull distribution shape the Weibull probability density function (PDF) to describe one of the three reliability failure regimes, namely, early defect driven fails, random fails, and wearout fails.

Both the Log-Normal and the Weibull distributions were used to describe TDDB failures. However, TDDB failure times are independent variables that are consistent with the Weibull distribution [41]. This is in agreement with the physical nature of

the breakdown, since the entire structure fails because of the failure of a localized region.

The physical origin of the breakdown is an active topic of debate and research. Competing models seems to describe breakdown equally well. These models can be broadly categorized into models of breakdown time as a function of electric field E. The exact of nature of dependence, along with the underlying physics, is still ambiguous. The most popular models describe breakdown as bond breakage due to thermo-chemical heating under stress, the E model, and damage incurred because of Cu diffusion, the $\sqrt{E}$ model. The details of the the E model and the $\sqrt{E}$ model are given below.

E Model: According to the E model, TDDB is due to field-enhanced bond breakage caused by the weakening of the polar bond, because of being stretched by the electric field, thereby making them susceptible to breakdown by a standard thermal process. The field dependence of the reduction of activation energy of bond breakage causes the degradation rate to increase exponentially with the field. Breakdown occurs when a localized density of broken bond causes the anode and the cathode to short [42–45]. The time-to-failure (TF) is given by

$$TF = A_\circ \exp\left(-\gamma E\right) \exp\left(\frac{Q}{k_B T}\right), \tag{1.2}$$

where γ is the field acceleration parameter, Q is the field dependent activation energy, $A_\circ$ is a material and process dependent constant and k_B is the Boltzmann constant. This model was developed to describe device dielectric breakdown but has been shown to fit well to backend of the line (BEOL) dielectric breakdown data [46].

$\sqrt{E}$ Model: The $\sqrt{E}$ model, pronounced "root-E" model, was specifically developed for backend dielectric breakdown. Low-k materials show two major leakage conduction mechanisms under high electric fields, namely, Poole-Frenkel emission (PF), and Schottky emission (SE), both showing the same $\sqrt{E}$ dependence of the current density. According to the $\sqrt{E}$ model, SE dominates at the lower bias region and PF

dominates at the higher bias region [47]. Specifically, under high field electrons, injected from the cathode and accelerated towards the anode, follow either SE or PF conduction at the interface of the dielectric and cap layer. Some of the electrons while being transported across the dielectric thermalize, a fraction of these thermalized electrons impact Cu atoms in the anode and accelerate the generation of positive Cu ions that inject into the dielectric along a fast diffusion path and create damage, leading to ultimate breakdown [47–50]. According to the $\sqrt{E}$ model

$$TF \propto \exp\left(-\sqrt{E}\right). \tag{1.3}$$

There are also other models of breakdown that relate electric field to TF [51–53].

It is prudent to mention here that the breakdown is a single event brought about by the random creation of defects, local conduction paths of very small dimensions [28]. The degradation is uniform across the dielectric. To link the random nature of the breakdown and its relation to defect generation, breakdown was shown to be caused by the generation of a critical number of defects within a cell when the dielectric area was partitioned into cells of small size [54]. Degraeve et al. proposed the "percolation model" in which the breakdown is triggered by a percolation path of defects, i.e. when the defects overlap to short the two electrodes [55]. Comparable results were shown using a similar approach in [56]. The percolation model successfully explains the dimensional dependance of dielectric breakdown, linespace in the case of backend dielectrics. Figure 1.1 shows the main idea of the percolation model.

1.1.1.5 Factors Impacting Dielectric Breakdown in Cu/Low-k Interconnects

Until the mid 1990s, research in dielectric breakdown was motivated by reliability of the gate dielectric. However, aggressive pursuance of Moore's Law meant that process engineers look for innovative solutions to overcome hurdles encountered in other parts of the chip. This led to the adoption of Cu as the metal of choice for interconnects, and low-k materials as ILD. Cu/low-k interconnects resolve the problem of RC delay

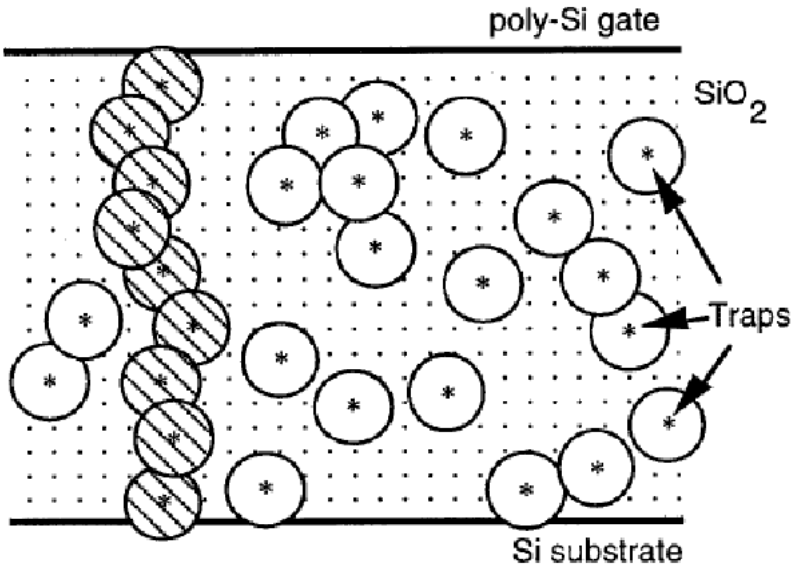


Figure 1.1: Illustration of the percolation model. The spheres are defects in the dielectric, breakdown takes place when defects overlap to form a conductive path between the two electrodes. Taken from [55].

but raise other questions, the most important among those being the reliability of these interconnect systems.

Historically, ILD reliability was not a concern because of the thickness of the dielectric. However, copper/low-k interconnect systems are vulnerable to TDDB because of the lower breakdown field strengths of porous low-k materials, the susceptibility of low-k materials to mechanical damage by CMP, and the high susceptibility of low-k materials to copper drift. These problems are compounded because the supply voltage is not scaled as aggressively as feature size, resulting in exponentially escalating electric fields among interconnects each technology generation. Porosity degrades the electrical and structural properties of copper/low-k systems further, because of, for example, the absorption of chemicals through pores that have an open connection to the surface.

Backend dielectrics are different than thin gate oxides in a number of ways. First, unlike the gate oxide, the backend dielectric undergoes many process steps that can

potentially damage the interfaces, which can become trap sites and assist in conduction. Second, the quality of the backend dielectric, which is deposited rather than thermally grown, is much poorer, resulting in higher defect densities. Third, backend geometries in chips include a wide variety of geometries, all of which may impact chip lifetime. The appropriate features and failure rates to extract and measure on a chip are not known. Moreover, due to the complexity of the structure, TDDB of copper/low-k damascene structures has to be assessed as a system of a dielectric, diffusion barrier, cap layer, and copper interconnect.

Since the electric field is much stronger at the corners and tips, feature geometries have to be taken into account when studying reliability concerns related to TDDB [57]. Porosity will continue to play an increasing role as the push for interconnects with lower values of k become stronger. An increase in porosity accelerates charge transport by the enhancement of electric field by the pores [58]. Low stiffness values of low-k materials can be improved by a strong Si-O covalent bond. Simulations have shown the presence of multiple atomic scaled cracks within the material [59]. Liners also affect the breakdown performance [60, 61]. The field acceleration parameter, γ, has been shown to be independent of the effective dielectric constant values [62]. Variation in linespace (S) may lead to deviations in observed values of γ, although γ is independent of S. Variation in S has emerged as a major concern and adversely impacts the lifetime of Cu/low-k interconnects, with smaller S leading to faster breakdown time [57, 63, 64]. Line topology is also an important concern; however stress conditions must be carefully selected to avoid artificial changes to distribution parameters [65]. It has been proposed that reduction in lifetime at smaller linespace may not be because of the spacing effect [66]. Vias also affect the TF adversely [67]. Unlike EM and bias temperature instabilities (BTI), TDDB performance is not affected by alternating current (AC) stress and the lifetimes are greater simply because of the lower effective stress [68]. CMP continues to be a source of early

TDDB fails [69] and also dictates wearout performance [47].

The aforementioned issues add to the complexity of the issue of reliability of Cu/low-k interconnect and hence call for a zero-defect manufacturing environment [47].

1.1.2 Variation in CMOS

Die-to-die shift in device performance was shown to be sufficiently modeled by the worst case design methodology [70]. However, because of shrinking dimensions, denser integration, and the increase in size of chips, within-wafer variation has become an important player in the overall operation of the circuit. Within-wafer variation can be further categorized into die-to-die variation and within-die variation, with the former being independent of design and systematic in nature, for instance when a likely cause is a process gradient, and the latter being more pronounced and dependent on design implementation [71].

Threshold voltage (V_{th}), which is the voltage at the onset of strong inversion, determines the gate bias that is needed to switch the transistor. However, owing to within-die variation, the value of V_{th} for two neighboring transistors on the same die can be different. The value of V_{th} can be determined from a given dataset by using a variety of techniques, the most common being the linear extrapolation method. V_{th} is taken as the voltage obtained after extrapolating the linear region of current characteristics, with respect to the gate-source voltage (V_{gs}), to zero drain current [72]. This method is sensitive to mobility degradation. Another method, called the transconductance (g_m) method, or maximum g_m method, gives V_{th} as the voltage at which the derivative of g_m, the second derivative of drain current (I_d) with respect to V_{gs}, is maximum [73]. This method was developed to avoid the dependence of the extracted threshold voltage on drain and source series resistances. However, it is sensitive to noise in the measurements. The constant current method is attractive

because of its simplicity. The gate voltage applied to give a certain predefined value of the I_d is taken as the V_{th} [74].

The statistical variation of V_{th} has become one of the major concerns associated with CMOS scaling. For accurate operation of circuits, accurate values of electrical parameters are required. This has placed great importance on the accurate prediction of statistical distributions of threshold voltage [75]. Devices have scaled to a point where there is a large variation in threshold voltage due to the number and location of dopants in the channel [76]. This has an adverse and direct effect on performance. For instance, in static random access memories (SRAM), random dopant fluctuations degrade the stability in operation [77]. Moreover, random dopant mismatch also impacts analog circuits that rely on device matching for their operation [78]. The standard deviation of variation in V_{th} can be described in terms of transistor dimensions [79], and is inversely proportional to transistor area.

MOSFET behavior is directly determined by the effective channel length, L_{eff}, where the inversion free carriers can be controlled by the gate voltage. However, the precise determination of L_{eff} has proved cumbersome, although a number of studies have concentrated their efforts on its extraction. Owing to their simplicity, the methods developed independently by Terada et al. and Chern et al. have been the most popular [80,81]. They are built on the idea that the channel resistance when plotted against the metallurgical length of the gate gives a straight line. However, this method may not give accurate results for deep sub-micron and nanometer devices. A modification to this method also fails because of the dependence of L_{eff} on gate voltage [82]. Taur et al. proposed a shift and ratio method, which uses a different definition of channel resistance to find the value of L_{eff} [83]. The conductance method for channel length extraction accounts for velocity saturation effects and can be used to extract L_{eff} at room temperatures as well as at very low temperatures [84]. Fikry et al. proposed a similar method [85]. Non-linear optimization methods have also

been used to extract L_{eff} [86,87] but they can give non-physical values and require a long computation time. C-V characteristics can also be used to extract L_{eff}, although they are not readily available in most cases [88–90].

L_{eff} is the smallest feature implemented on silicon; thus it shows the largest within-die variation due to optical proximity and other process related effects [91]. With scaling, the variability in critical dimension (CD) will increase. However, due to short channel effects, the dependency of drain current on L_{eff} will decrease. Consequently there will be a reduction in performance variation from L_{eff} [92]. Nevertheless, the International Technology Roadmap for Semiconductors (ITRS) lists the control of CD as one of the challenges for manufacturing [93].

The efforts of addressing variability can be classified into statistical metrology, advanced process control and advanced equipment control, and design for manufacturability (DFM) [94]. Variability and the continuing trend of scaling has increased the importance of accurate and simple analytical models. However, this task has been complicated by the quantum mechanical effects in nanometer devices [95].

The term "compact model" is defined as an analytical model comprising sequences of equations that originate from physical considerations but are subsequently modified to achieve a better fit to measured data [96]. Compact models can be either surface-potential based [97,98] or inversion-charge based [99,100]. Although all these models involve empirical fitting to a varying extent, they can also lead to inaccurate model sensitivity analysis [101]. Furthermore, variation in parameters can add uncertainty to these models, calling for a need to accurately capture the process variation. This has led to the steady growth of the field of statistical metrology, the study of how to characterize and model process and environmental variations [94,102]. Traditionally, variability has been modeled by detailed circuit simulations, but increasing integration densities prohibit such an approach because of the long computation time [103]. The fact that within-die variation is spatially correlated adds to the problem [104]

particularly for gate length variation [105]. However, lack of spatial correlation has been shown in V_{th} variation [106]. Test structures designed specifically for the purpose of measuring and modeling within-die variation have also been used [107–111]. This can aid in linking statistical metrology and circuit design by mapping process variation onto designer-controlled variables through modeling [112].

1.2 Research Objective

The objective of the research is two-fold. Firstly, at the back-end, the goal of the research is to identify critical physical design features affecting backend low-k TDDB. The research also notes that accurate analysis of empirical data requires accurate extraction of failure distribution parameters. Thus, methods of determining lifetime requirements in the presence of uncertainty in distribution parameters are developed. The impact of critical physical design features for backend low-k TDDB is unified in a framework to predict backend low-k TDDB chip reliability. Secondly, at the front end, the research aims to model variation in electrical and physical device parameters to aid in circuit design and analysis. The ultimate goal of the research is to introduce reliability and manufacturability in chips by accurately predicting chip behavior. Ultimately this will help in decreasing costs through better yield, and increased lifetime through more reliable designs.

1.3 Overview of the Thesis

1.3.1 Composition

The thesis has been partitioned into two sections. Interconnect reliability, specifically backend low-k TDDB, forms the first portion and device variability forms the second.

1.3.1.1 Reliability

Chapter 2 looks at interconnect geometries to determine their potential impact on failure rates. Failure rates vary both as a function of Weibull statistics and as a

function of die-to-die linewidth variation. Determination of whether or not a process satisfies lifetime requirements should take both factors into account . As variation becomes large, the lifetimes that achieve the same target probability of failure are orders of magnitude lower than without die-to-die linewidth variation.

Chapter 3 shows that low-k TDDB may vary as a function of metal linewidth, when the distance between the lines is constant. Modeling requires determining the relationship between TDDB and layout geometries. Models are computed to estimate TDDB as a function of linewidth, and the cause of variation in TDDB behavior is investigated.

Chapter 4 proposes a framework to analyze circuit layout geometries to predict chip lifetime due to low-k time-dependent dielectric breakdown. The methodology will use data from test structures, which have been designed and fabricated to detect the impact of area and metal linewidth on low-k TDDB, as inputs.

1.3.1.2 Variability

In Chapter 5, a methodology is developed to characterize within-die variation in transistors for use in the development of standard cell models and to enable the incorporation of both random and systematic variation in circuit analysis. Within-die variation in the manufacturing process results in variation in device and circuit characteristics. It can range from purely random to purely systematic. A methodology to determine the impact of within-die variation on circuit timing, separately from die-to-die variation, is proposed.

Chapter 6 concludes the thesis with a summary of the research and ideas for future work.

PART I

RELIABILITY

CHAPTER II

A METHODOLOGY TO EXTRACT FAILURE RATES FOR LOW-K DIELECTRIC BREAKDOWN WITH MULTIPLE GEOMETRIES

2.1 Introduction and Motivation

This chapter explores the impact of interconnect area and geometry on backend Cu/low-k TDDB failure rates and develops failure-rate models for determining the time-to-fail at small percentiles (the tails of the distribution) while accounting for the observed curvature in Weibull plots. Data obtained from back-end dielectric breakdown lifetime measurements is directly used to determine the impact of linewidth variation. The use of this information to determine equivalent lifetime requirements— that consider both the impact of die-to-die linewidth variation and the traditional failure rate distribution, modeled by Weibull statistics— is also discussed.

Measuring backend dielectric reliability relies on comb structures. A voltage difference is applied to the comb structure, which creates a lateral electric field through the intra-layer dielectric between the fingers of the comb, which are separated by the minimum space design rule. The current between the fingers of the comb is monitored. Breakdown is observed when the current exceeds a fixed threshold.

Data are collected for a sample of comb structures. In collecting data for a sample of comb structures, the data is ordered from the shortest to the longest breakdown time. Each time point is assigned a probability point, P, by partitioning the probability scale equally. An example is shown in Figure 2.1. The data is fit by a distribution, either the Weibull distribution or the Log-normal distribution, in order to enable extrapolations to lifetimes at low percentiles. Lets consider the Weibull distribution,

as an example. When constructing a model by fitting a Weibull distribution to a data set, two parameters are extracted: the characteristic lifetime (63.2% probability point), η, and the shape parameter, β. The resulting data is then scaled to use conditions and the vulnerable area corresponding to the chip.

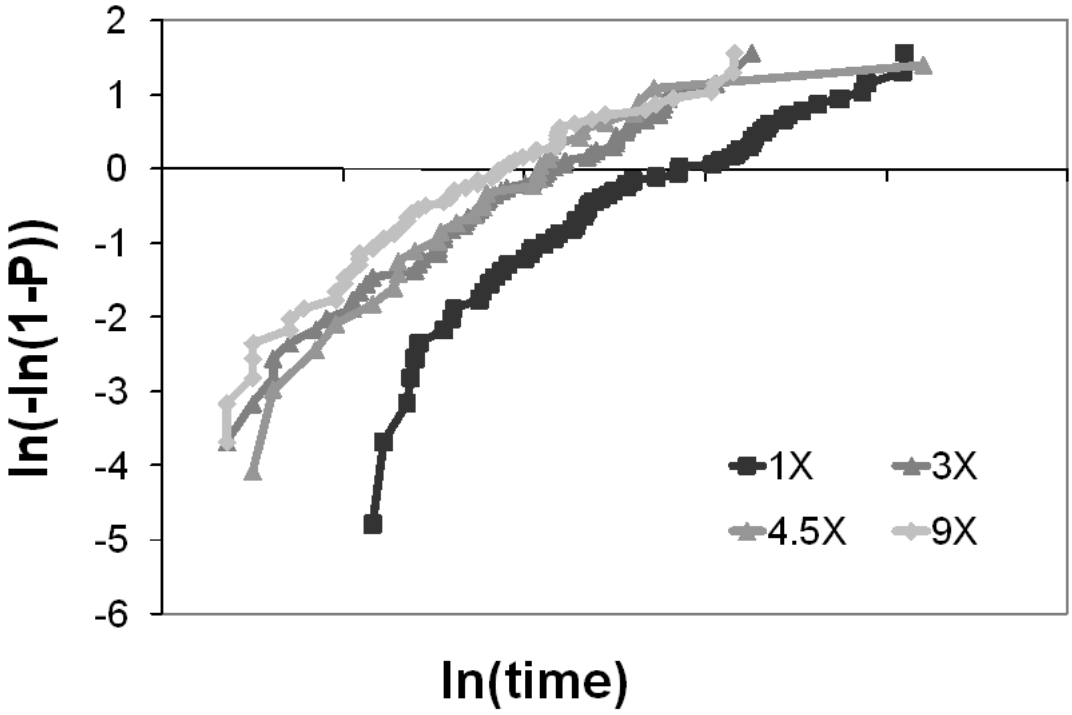


Figure 2.1: Example Weibull plot of $\ln(-\ln(1-P_i))$ vs. $\ln(\text{time-to-failure}(i))$ for comb test structures with four areas: $1X, 3X, 4.5X$ and $9X$.

This chapter has two objectives. First, to investigate the impact of multiple geometries on failure rates. Second, to build models of the failure rate as a function of relevant geometry parameters to determine failure rates at small percentiles.

When considering multiple geometries, it was noted that field enhancement occurs at the tips of the comb structures and in the presence of misaligned vias. Prior work has indicated that failure sites correspond with locations where there is field enhancement [57, 113]. In order to address this concern, a set of test structures, that can separate the impact of field enhancement at the comb tips and as a function of

the number of vias, was designed. Analysis of data from these test structures required developing a methodology to determine failure rates for structures containing multiple geometries. Determination of failure rates at low percentiles requires constructing a model. As can be seen from Figure 2.1, the data points do not fall on a straight line, as expected for Weibull statistics. In [114] it was noted that linewidth variation can be as large as $\pm 30\%$. This variation distorts the Weibull curves used to determine the lifetime of a structure. As a result, direct extraction of the parameters, η and β, is inaccurate.

It is suggested in [114] that die-to-die linewidth variation can be eliminated during Weibull parameter extraction through calibration of lifetime measurements based on capacitance measurements, from which the mean space between the lines of the comb structure can be computed. However, because of the complexity of the structure, capacitance is also impacted by variation in the low-k dielectric constant as a function of the composition of the dielectric stack, and consequently, variation in the dielectric constant is confounded with variation in distance. Hence, capacitance measurements overestimate linewidth variation.

This chapter uses data from backend dielectric breakdown lifetime measurements directly to determine the linewidth variation, and then uses this information to take into account both the impact of die-to-die linewidth variation and the traditional failure rate distribution, modeled by Weibull statistics, when computing lifetime at a specified percentile.

Note that in this chapter, as well as the proceeding chapters, the term defect density is used. Defect density alludes to an extrinsic distribution due to defects in the dielectric, however wearout distributions are intrinsic distributions. As pointed out in [28], a more appropriate term for dielectric breakdown is event density since that would represent the statistical nature of dielectric breakdown. However, in this document the more general term defect density is used because of its prevalence in

literature. Thus, the term "defect density" should not be taken to imply that there are extrinsic defects in the dielectric.

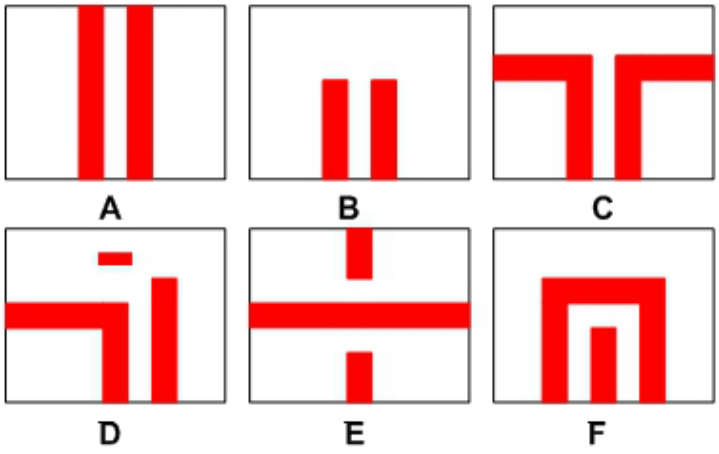
This chapter is organized as follows. In the next section, previous data, which indicated the potential impact of field enhancement on low-k dielectric lifetimes, is discussed together with the test structures that were designed to distinguish the impact of area and field enhancement from the tips of combs. Section 2.3 presents the methodology to extract the impact of field enhancement separately from breakdown due to area only, and analyzes empirical data that was collected from the test structures. Section 2.4 presents a methodology to extract die-to-die linewidth variation directly from the failure rate data, since die-to-die linewidth variation provides a correction in the slope of the failure rate distribution at the characteristic lifetime. Section 2.5 demonstrates the use of the Weibull failure rate model in combination with the extracted die-to-die linewidth variation to determine probabilities of failure as a function of lifetime, combining both sources of degradation. Section 2.6 concludes the chapter with a summary.

2.2 Determination of the Impact of Field Enhancement

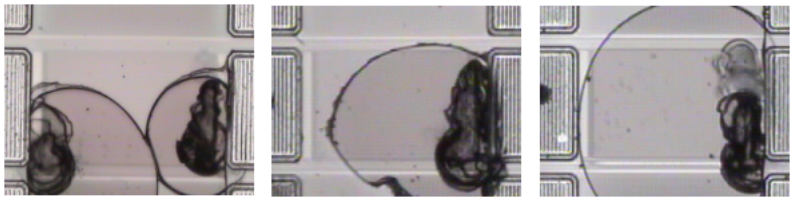
2.2.1 The Role of Field Enhancement in Breakdown

It is generally assumed that the vulnerable area in a comb test structure is solely a function of the area and distance between the fingers of the comb. However, the electric field between the fingers of the comb is non-uniform, and there is significant field enhancement at the comb tips that must be taken into account because the breakdown is field driven.

In order to analyze backend geometries, several circuits were synthesized using standard placement and routing tools from which the most frequent patterns in the backend geometries were extracted [57]. Besides parallel lines, the five other patterns in Figure 2.2(a) were frequent. The results were consistent among the circuits.



(a)



(b)

Figure 2.2: (a) Frequent interconnect geometries in sample circuits. (b) Example failure sites after breakdown. The dark areas show the failure sites, which are near the pads and at the tips of the combs.

For each of these structures, a 3D finite element model was constructed to determine the electric field distribution. The field was found to be highly non-uniform, with peak fields from some geometries, such as Geometry F, exceeding that of the parallel line structure by a factor of $2-3$ [57], because of the bends and corners. Also, high fields could be found at the top (cap layer) and bottom interfaces for all structures [57], because of the sharp edges. These high fields at the top, cap layer interface, are especially problematic, since this interface is formed by CMP. This is a low quality interface, which contains many dangling bonds, facilitating copper ion drift or the formation of a percolation path.

It should be noted that the comb test structure has two types of geometry: parallel lines with minimum space and tips. Field enhancement can potentially occur at the

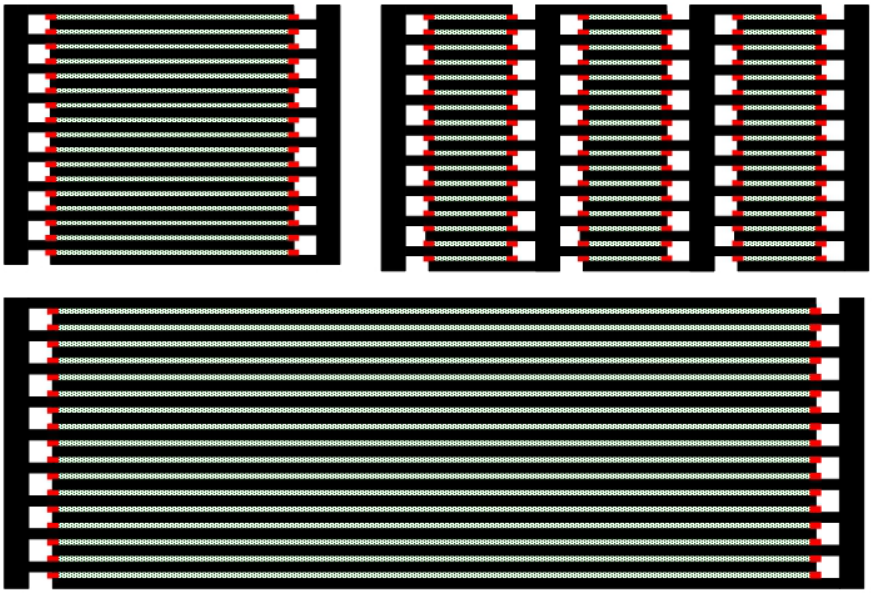
tips. Experimental results for $0.18\mu m$ technology, involving stressing a set of industrial comb structures made with copper and low-k materials, indicated the potential role of field enhancement. In fact, even though the test structures had long parallel lines with minimum space, most failure sites coincided with peaks in electric field at the corners in the structure. Examples are shown in Figure 2.2(b). This is similar to results in [113], where failure sites also coincided with electric field enhancement.

2.2.2 Test Structures and Data Collection

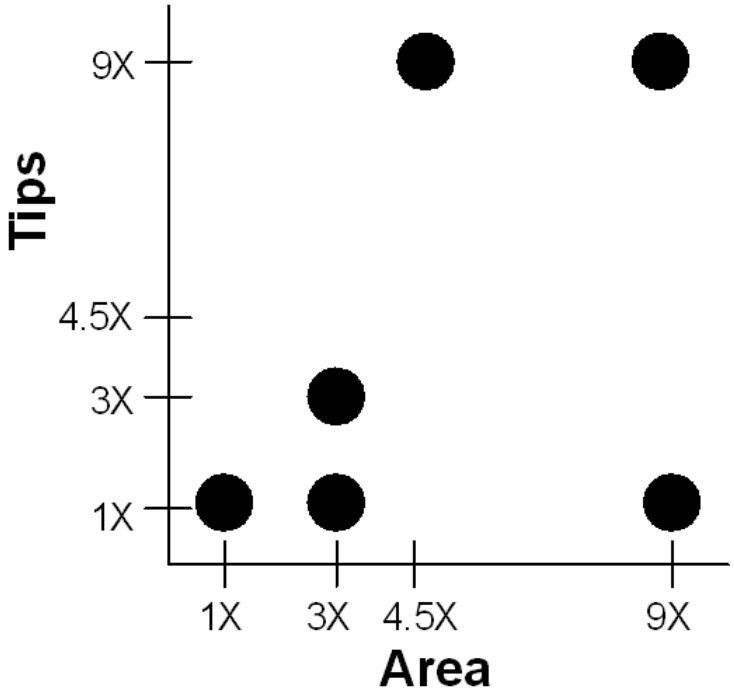
Clearly, it is necessary to isolate the impact of the vulnerable area between parallel lines and tips in experimental work. Figure 2.3(a) illustrates the definition of parallel line capacitors and tips used to isolate failure rates. Data from the test structures will be analyzed in pairs, where one of the test structure pairs holds area constant and varies tips, while another pair holds the number of tips constant and varies area. The matrix of test structure features that are used in this chapter is shown in Figure 2.3(b). Data was collected by applying a voltage difference of $4V$ at 125°C.

2.2.2.1 Test Results

The measurement data related to area with fixed $1X$ tips is shown in Figure 2.4(a). The measurement data related to tips with fixed $3X$ area and $9X$ area are shown in Figures 2.4(b) and 2.4(c), respectively. The results indicate a strong impact of area. The data on tips is inconclusive. A comparison of the $3X$ area structures indicates an impact of tips, while a comparison of the $9X$ area structures shows no impact of tips. The last data point with $4.5X$ area needs to be considered to determine the impact of tips. However, to use this dataset, it is necessary to first analyze the impact of area so that the other datasets can be modified to correspond to the $4.5X$ area points.

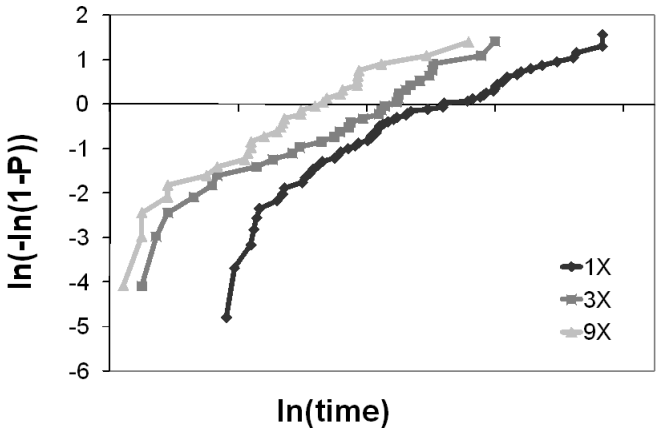


(a)

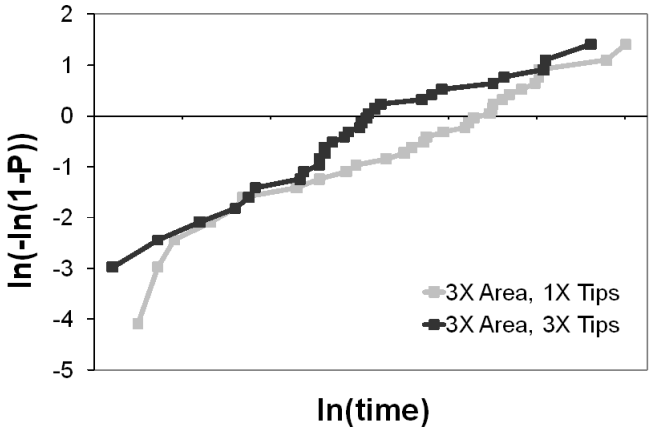


(b)

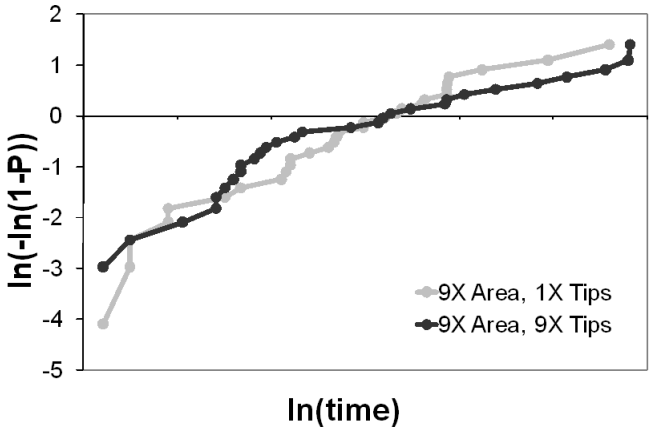
Figure 2.3: (a)Comb test structure to measure backend dielectric breakdown. The vulnerable area is shaded in green. The vulnerable tips are shaded in red. (b) Test structures set to extract and verify the failure rates due to tips and area. The dots denote the availability of test structures with the corresponding area and number of tips.



(a)



(b)



(c)

Figure 2.4: Measurement data for (a) test structures with $1X$ tips and $1X$, $3X$, and $9X$ area, (b) test structures with $3X$ area and $1X$ and $3X$ tips, and (c) test structures with $9X$ area and $1X$ and $9X$ tips.

2.2.3 Analysis Methodology

Suppose that there are two independent failure mechanisms affecting a test structure, each associated with a defect density, $\lambda_1(t)$ and $\lambda_2(t)$, which are functions of time. Each of these defect densities is associated with feature areas, A_1 and A_2. In the case of tips, the feature area is the number of tips. Using the Poisson model, the probability of survival for a structure containing feature i is $R(t) = \exp(-\lambda_i(t)A_i)$. If the failure mechanisms are independent, then the joint probability of failure, $P(t) = 1 - R(t)$, is

$$P(t) = 1 - \exp(-\lambda_1(t)A_1)\exp(-\lambda_2(t)A_2). \tag{2.1}$$

The standard statistical function to characterize TDDB reliability failure distributions is the Weibull distribution:

$$P(t) = 1 - \exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right). \tag{2.2}$$

In order to extract distribution parameters, one plots $\ln(t)$ vs. $\ln(-\ln(1 - P(t)))$, since (2.2) can be rewritten as

$$\ln(t) = \ln(\eta) + \frac{1}{\beta}\ln(-\ln(1 - P(t))),$$

or

$$\ln(-\ln(1 - P(t))) = \beta(\ln t - \ln \eta). \tag{2.3}$$

It should be noted that if there are multiple failure mechanisms, then by rearranging (2.1), we have

$$\ln(-\ln(1 - P(t))) = \ln\left(\sum_i \lambda_i(t)A_i\right). \tag{2.4}$$

Hence, it can be seen that the Weibull distribution plot also gives an indication of the probability distribution function for the number of defects at breakdown, i.e., $\lambda_i A_i$.

Suppose that two test structures contain all the same features, except in one structure there is $3X$ of a target feature vs. $1X$ in the other structure. The difference

between these two structures is $2X$ of the target feature. The number of defects in $2X$ of the target feature is

$$\lambda_{2X}(t)A_{2X} = \lambda_{3X}(t)A_{3X} - \lambda_{1X}(t)A_{1X}. \tag{2.5}$$

Therefore values of $\lambda_{2X}(t)A_{2X}$ can be extracted, using the Weibull distribution plots for the $3X$ and $1X$ test structures. Using Equation (2.4), at any time, t^*, the failure distribution for $2X$ of the target feature is

$$\ln(-\ln(1 - P_{2X}(t^*))) = \ln(\ln(1 - P_{1X}(t^*)) - \ln(1 - P_{3X}(t^*))), \tag{2.6}$$

where the values for $P_{1X}(t^*)$ and $P_{3X}(t^*)$ are known from the measured data from the $1X$ and $3X$ test structure. From Equation (2.6) we can solve for $P_{2X}(t^*)$.

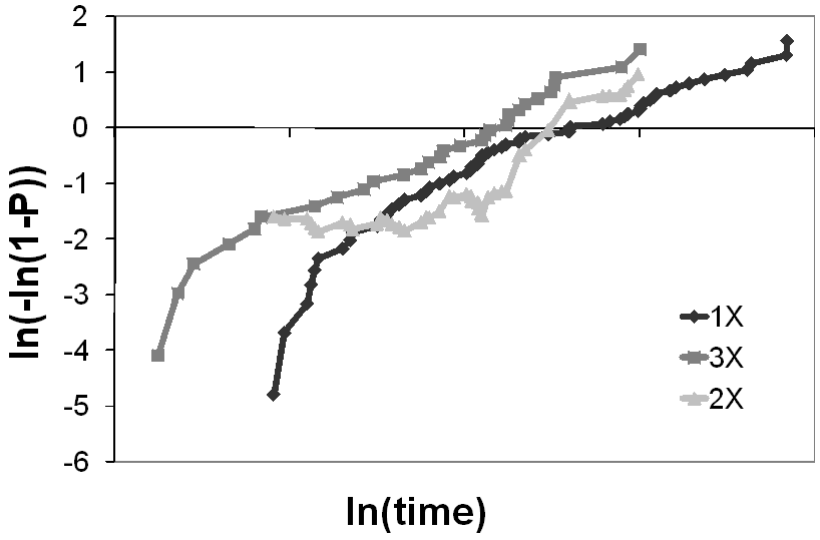
2.2.4 Separation of the Impact of Field Enhancement at Tips

In Figure 2.3(b) the row associated with $1X$ tips is used to extract the failure rate due to area between parallel lines with minimum space. The analysis methodology subtracts the impact of $1X$ tips from the measured results for these test structures, in accordance with (2.6).

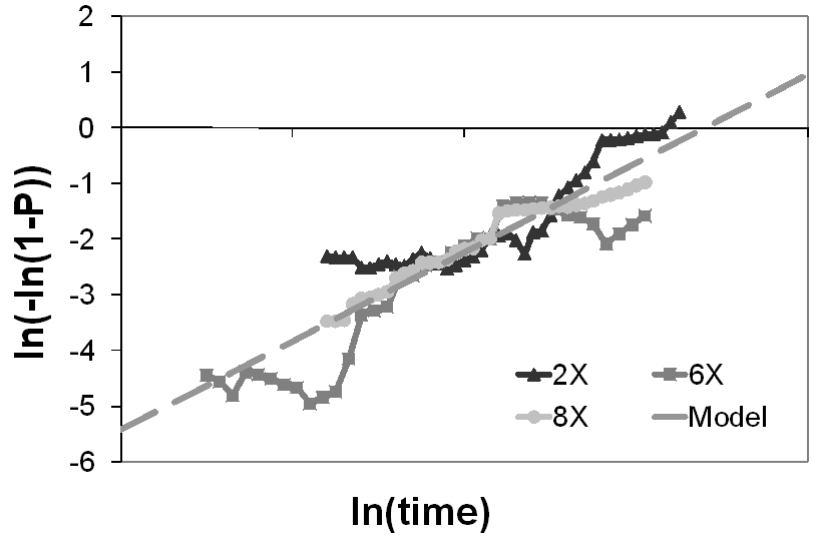
Figure 2.5(a) shows extracted values for $2X$ of the target feature (area), using $1X$ and $3X$ test structures. It plots $\ln\left(\sum_i \lambda_i(t)A_i\right)$.

The resulting curves for $2X$ area are non-monotonic because of noise in the dataset. Similarly, datasets for $6X$ area, using $9X$ and $3X$ test structures, and for $8X$ area, using $9X$ and $1X$ test structures, were extracted. These datasets are merged prior to computing a model, with a transformation of the probability scale [115–117]. Specifically, $\ln\left(-\frac{1}{N}\ln(1 - P(t))\right)$ vs. $\ln(t)$ is plotted, where $N = 2, 6$, and 8 for the $2X$, $6X$, and $8X$ datasets, respectively. Figure 2.5(b) shows the merged datasets, together with the fitted Weibull model. When the datasets are merged, the resulting distributions fall on a straight line. The equation for $1X$ area is

$$\ln(t) = \ln \eta^* + \frac{1}{\beta^*}\ln(-\ln(1 - P(t))),$$



(a)



(b)

Figure 2.5: (a) Extracted failure distribution for $2X$ area. (b) Merged dataset for $2X$, $6X$, and $8X$ area, with the probability scale modified to correspond to $1X$ area. The plots are for $\ln\left(\sum_i \lambda_i(t)A_i\right)$, but labeled as $\ln(-\ln(1-P))$, since they are equivalent according to (2.4).

or

$$\ln(-\ln(1-P(t))) = \beta^*(\ln t - \ln \eta^*), \tag{2.7}$$

where the extracted Weibull characteristic lifetime is η^* and the extracted Weibull shape parameter is β^*.

The measurement data related to tips, with fixed $3X$ area and $9X$ area is inconclusive. A comparison of the $3X$ area structures indicates an impact of tips, while a comparison of the $9X$ area structures indicates no impact of tips.

The extracted model is used to convert all datasets to correspond to failure rates with $9X$ area with varying numbers of tips. To do this, the area model is used to add defect densities. In other words, for the $1X$ model, defect densities corresponding to $8X$ area are added. If η^* and β^* are the parameters corresponding to the $1X$ area model, and if $P_i(t_i)$ is the probability point corresponding to breakdown time, t_i, then the probability point, $\acute{P}(t_i)$, corresponding to $9X$ area at breakdown time, t_i, is

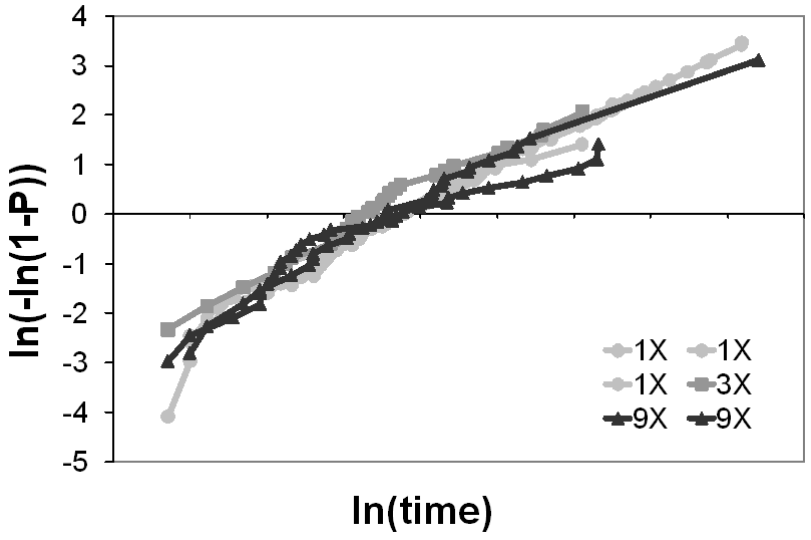
$$\ln(-\ln(1-\acute{P}(t_i))) = \ln\Big(-\ln(1-P_i(t_i)) + 8(\exp(\beta^*(\ln t_i - \ln \eta^*)))\Big). \tag{2.8}$$

The number of defects at failure for $8X$ area is $8(\exp(\beta(\ln t_i - \ln \eta^*)))$. Similarly, for $4.5X$ area, defect densities corresponding to $4.5X$ area are added. The results after converting all datasets to $9X$ area are shown in Figure 2.6(a). This figure shows *no impact of tips*. Variation among datasets appears to be random.

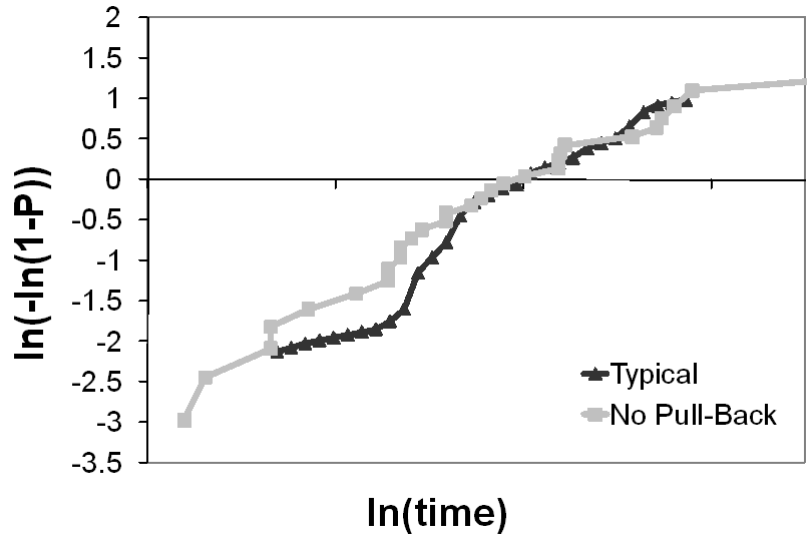
2.2.5 Detection of Field Enhancement from Geometry F

Section 2.2.1 indicated that the peak fields were significantly higher for Geometry F in Figure 2.2(a). Consequently, comb structures that contain the wrap around Geometry F at the tips of the combs were designed by eliminating the pull-back at the ends of the combs.

There is no test structure with area matching that of the test structure with Geometry F. The vulnerable area for the test structure with Geometry F is $1.4X$. Hence the methodology described in Section 2.2.3 to combine test structure data was



(a)



(b)

Figure 2.6: (a) Failure rate distributions for $9X$ area, with varying numbers of tips. The model in Figure 2.5(b) is used to convert all datasets to $9X$ area. $1X$, $3X$, and $9X$ refer to $1X$, $3X$, and $9X$ tips. (b) A comparison of lifetimes for test structures with and without pull-back at the tips of the combs, to determine vulnerability of Geometry F in Figure 2.2(a).

used to interpolate the test results from the test structures with $1X$ and $3X$ area, to generate data for $1.4X$ area. The data for test structures with and without pull-back are shown in Figure 2.6(b). It can be seen that the characteristic lifetime is not affected by Geometry F. However, the weaker population is impacted.

2.3 Failure Rate Modeling

This section discusses the use of results in the previous section for modeling the failure rate in the presence of die-to-die linewidth variation. The Weibull shape parameter is extracted first via area scaling and then the impact of die-to-die linewidth variation is extracted via the slope of the Weibull curve.

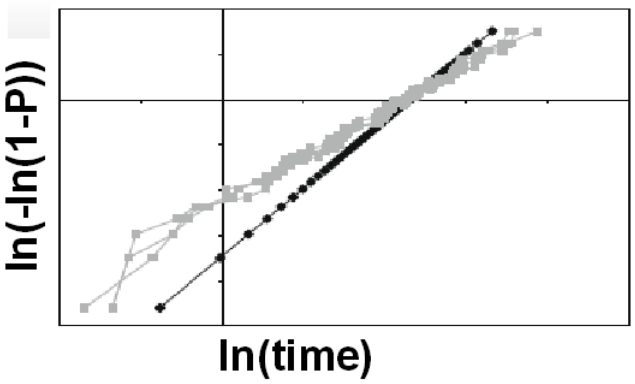
2.3.1 Extraction of the Weibull Shape Parameter via Area Scaling

In this study, test structures with four different areas: $1X$, $3X$, $4.5X$, and $9X$, as shown in Figure 2.3(b) and implemented with $45nm$ technology, were used. For this set of test structures, the distance between the lines of the comb, which determines the applied electric field through the dielectric, is fixed. In addition, a test structure with a smaller distance between the lines with $1X$ area and $1X$ tips was also used.

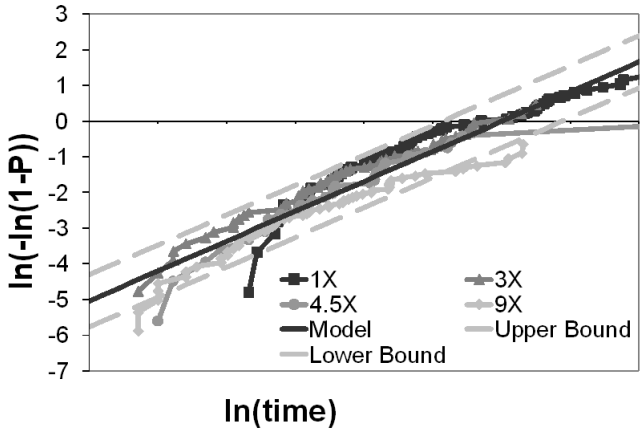
The data for the four different areas are shown in Figure 2.1. It can be seen that the failure rate distributions are not linear, as expected with the Weibull distribution model. The standard approach to determine the characteristic lifetime η and the Weibull shape parameter β is to transform the probability scale of all of the datasets by plotting $\ln(-\ln(1-P(t))/N)$ versus $\ln(t)$, where N is the area ratio and then fit a straight line to the data by regression, as shown in Figure 2.7(b).

Figure 2.7(a) shows the simulated distortion of the Weibull distribution due to die-to-die linewidth variation. It can be seen that variation creates curvature in the failure rate distributions. It also degrades the measured Weibull shape parameter, β, while η is less affected.

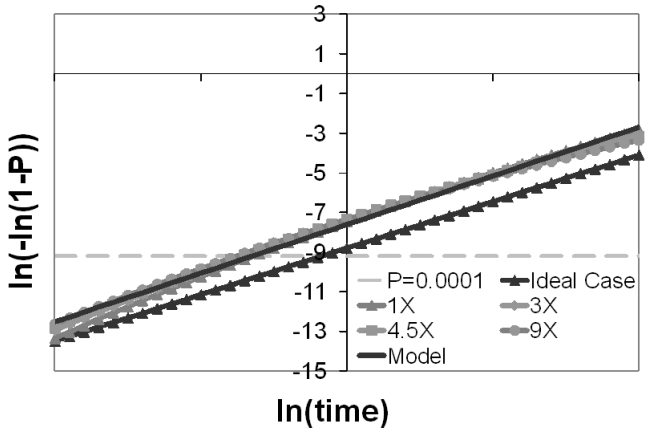
The effect of die-to-die linewidth variation is different than line edge roughness,



(a)



(b)



(c)

Figure 2.7: (a) Simulated Weibull distribution with and without die-to-die linewidth variation. The black line corresponds to the ideal case, while the gray lines are random samples impacted by die-to-die linewidth variation. (b) Model for $1X$ area for the merged data sets, combined with 90% confidence bounds. (c) Simulated Weibull distributions with and without die-to-die linewidth variation. The data with die-to-die linewidth variation for $1X$, $3X$, $4.5X$, and $9X$ area are merged using the method in [115–117] to correspond to $1X$ area and are modeled with a straight line.

which does not create curvature and degrade β, but causes a shift in η.

The extracted Weibull parameters are used to find the lifetime at a low probability point, such as the 0.0001 point. This is well beyond the data points in the dataset, since $\ln(-\ln(1-0.0001)) = -9.21$.

Errors in the extraction of the Weibull shape parameter, β, can create significant errors in the estimated lifetime at low probabilities. Consider a dataset with a true Weibull distribution, shown in Figure 2.7(c). If die-to-die linespace variation of 10% is introduced before merging the datasets for $1X$, $3X$, $4.5X$, and $9X$ area, it can be seen that die-to-die linewidth variation creates at least an order of magnitude error in the expected lifetime at the 0.0001 probability point.

Instead, as in [114], the area scaling formula is used to extract the Weibull shape parameter. The relationship between characteristic lifetimes for structures with different areas is given by

$$\ln \eta_{NX} = \ln \eta_{1X} + \left(\frac{1}{\beta}\right) \ln \left(\frac{A_{1X}}{A_{NX}}\right), \tag{2.9}$$

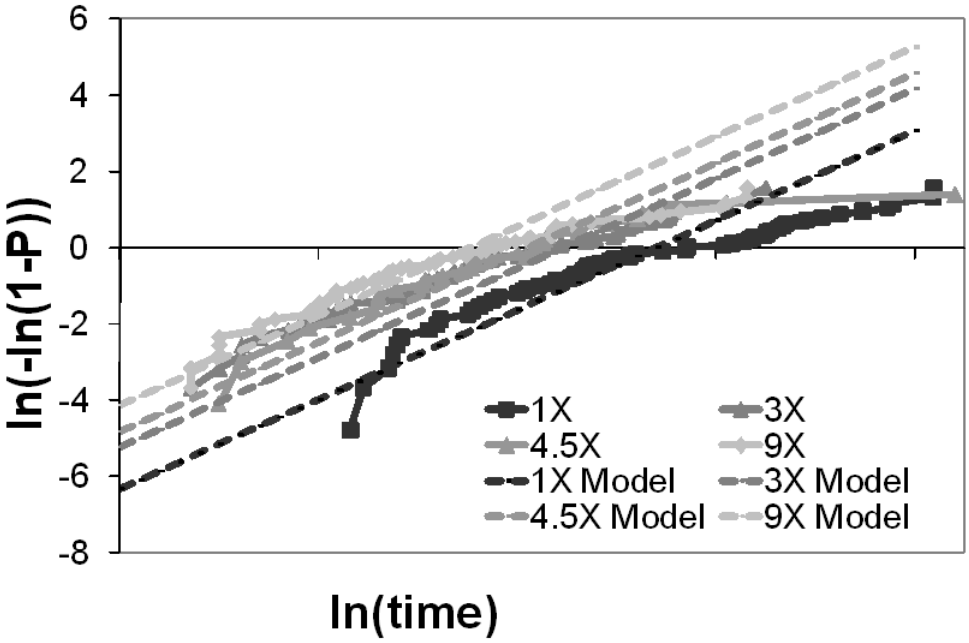
where η_{NX} and η_{1X} are characteristic lifetimes for the NX and $1X$ area test structures, respectively, and $\frac{A_{NX}}{A_{1X}} = N$ is the area ratio between the structures.

In order to extract the characteristic lifetime, for each area ratio, a quadratic model is fitted to each dataset to find the x-intercept. This reduces noise in the extraction of the characteristic lifetime.

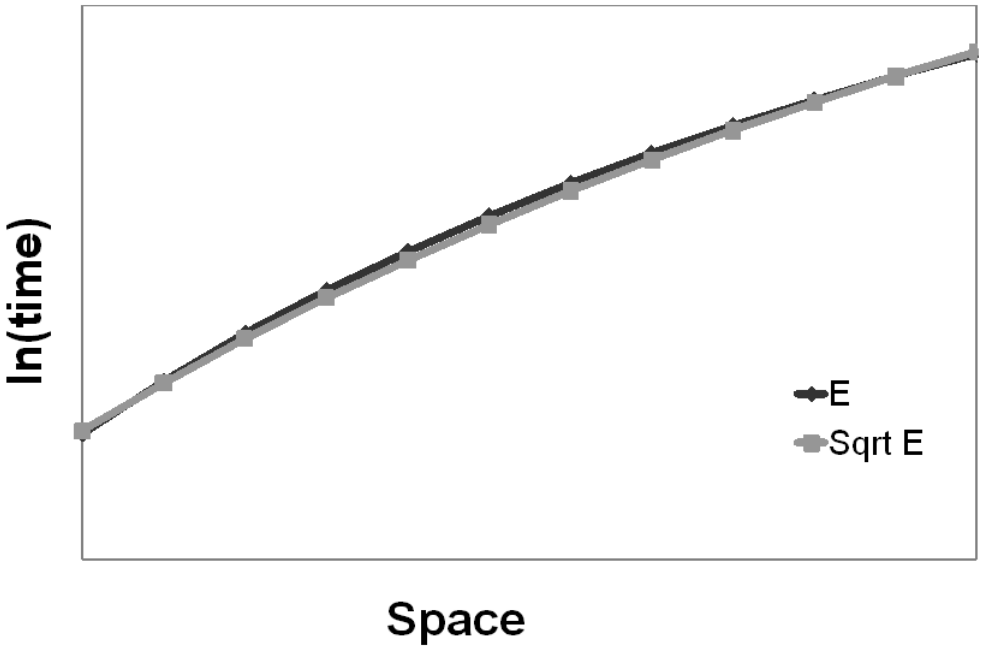
Figure 2.8(a) shows the extracted models, in comparison with the original datasets. Note that the slope does not match, due to die-to-die linewidth variation in the datasets.

2.4 Extraction of the Impact of Die-to-die Linewidth Variation via the Slope of the Weibull Curve

Figure 2.7(a) indicates that the impact of die-to-die linewidth variation is to degrade the slope of the Weibull curves. Therefore, the deviation in the slope from β was used



(a)



(b)

Figure 2.8: (a) Models for the datasets using the area scaling formula to find the Weibull shape parameter. The x-intercept is the characteristic lifetime determined with Equation (2.9). (b) Extracted relationship between characteristic lifetime and the distance between the lines for both the E and $\sqrt{E}$ models.

to find the die-to-die linewidth variation. However, the variation of characteristic lifetime as a function of distance between the lines must be known.

There are two dominant models of characteristic lifetime as a function of the distance between the lines. With the E model [46, 118]

$$\ln \eta = a_1 - \frac{b_1}{s}, \tag{2.10}$$

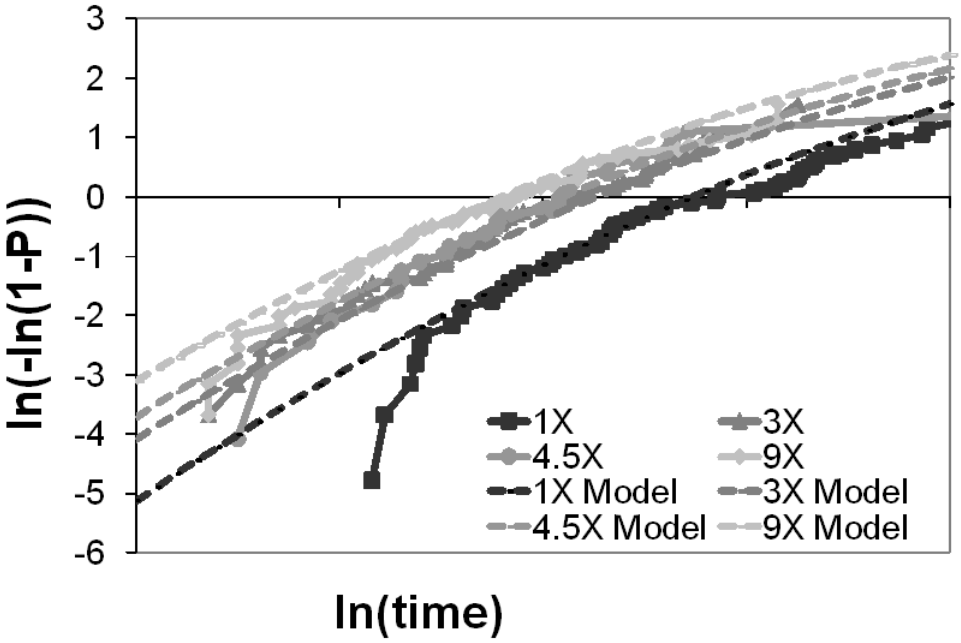
where s is the distance between the lines and a_1 and b_1 are fitting constants. With the $\sqrt{E}$ model [48–50]

$$\ln \eta = a_2 - \frac{b_2}{\sqrt{s}}, \tag{2.11}$$

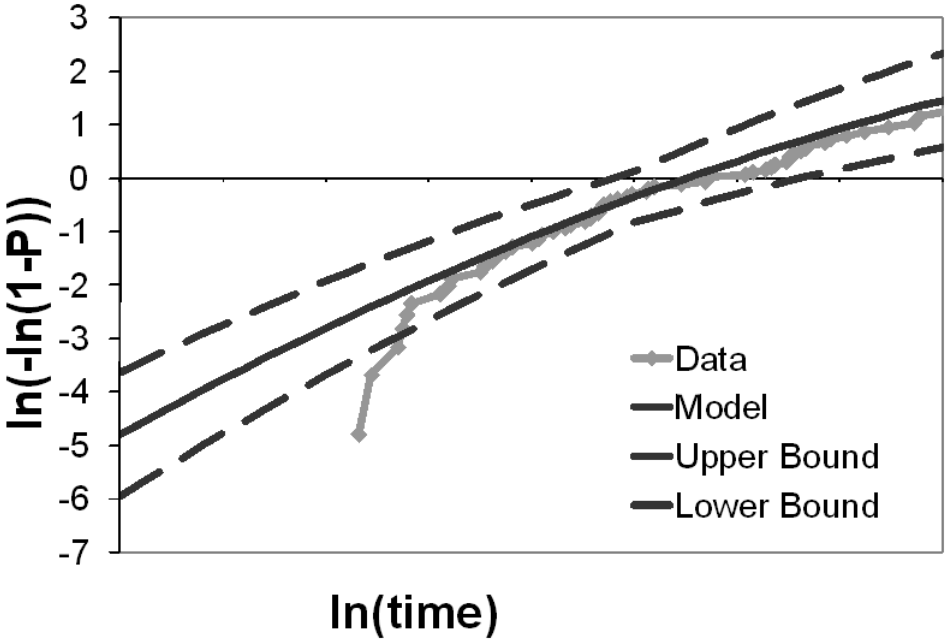
where a_2 and b_2 are fitting constants.

In order to determine the impact of line spacing on characteristic lifetime, dataset with a different distance between the lines was used. With two values for s, values of the constants were determined. The result is shown in Figure 2.8(b). It can be seen that the two models are almost identical for the range of data available from the two test structures.

However, in fitting the models in Equations (2.10) and (2.11), it is necessary to take into account any bias in printing, since the drawn distance between the lines in the layout may not match the printed distance exactly. Print bias is the difference between the distance between the lines in the layout and those in silicon. The manufacturer selects it to minimize interconnect delays. Hence, normally distributed die-to-die variation was assumed and the fit between the data and the model was optimized, given the value of β determined in Section 2.3.1 and an assumed value of print bias, to find the standard deviation of die-to-die variation. The results are shown in Figure 2.9(a). It can be seen that the models resemble the datasets. In addition, it was found that the extracted model is virtually identical for the E and $\sqrt{E}$ models and for any choice of print bias. In fact, for each value of bias, the optimum fit resulted in the same value for the standard deviation for the distance between the lines.



(a)



(b)

Figure 2.9: (a) Models for the datasets using the area scaling formula to find β, combined with a correction in the slope and curvature due to die-to-die variation. The standard deviation of die-to-die variation was determined by optimizing the fit between the model and the data. The $\sqrt{E}$ model was used to determine the variation in lifetime as a function of distance between the lines. (b) The extracted model for the $1X$ dataset and the 90% confidence bounds.

The confidence bounds for the $1X$ model are shown in Figure 2.9(b). Variation is dominated by accuracy in extraction of the Weibull parameters with Equation (2.9).

2.4.1 Variation in Failure Statistics due to Vias

Vias above and below interconnect lines can cause local field enhancement in the presence of alignment errors. Moreover, prior work indicated that there can be degradation in failure rate statistics due to the presence of vias [67, 119].

Test structures were implemented with 0, 2000, and $10,000$ vias below the lines in the comb structures. In order to compare the failure rate statistics, the model and confidence bounds computed in Section 2.4 were used for the test structure without vias with a correction to the characteristic lifetime due to a small change in the test conditions. The results are shown in Figure 2.10.

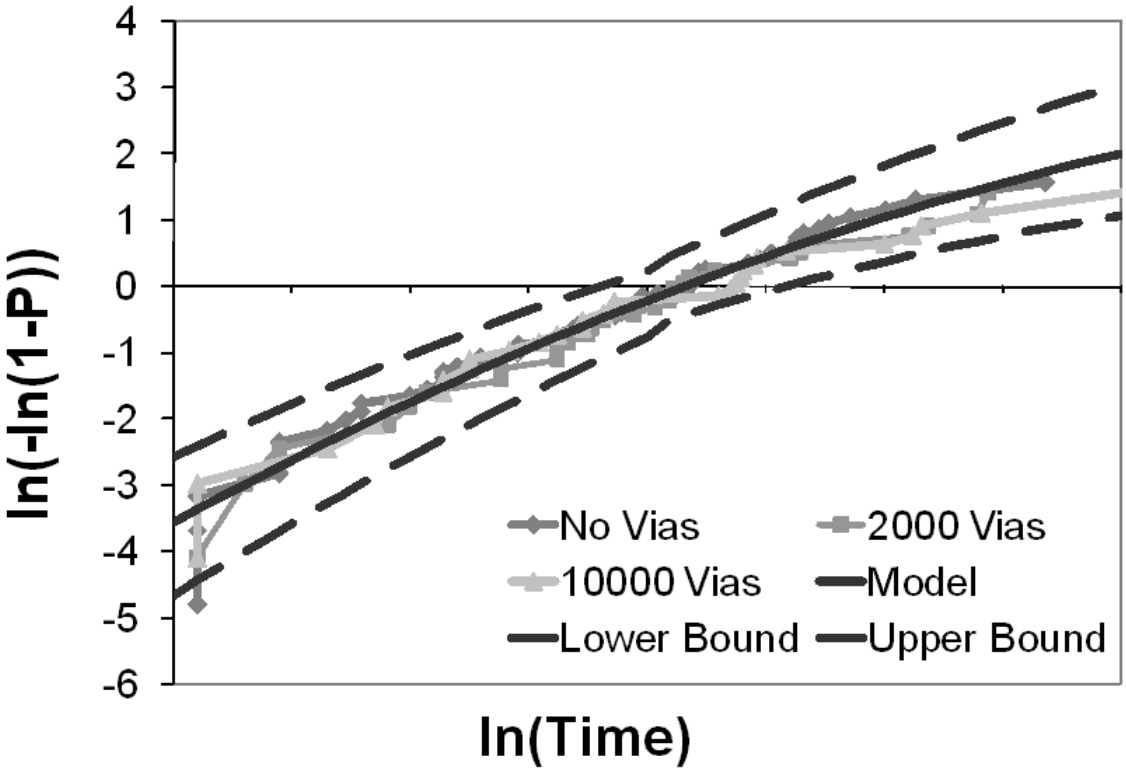


Figure 2.10: Comparison of the failure rate of test structures with and without vias, together with confidence bounds on the failure rate dataset without vias.

Data for test structures containing vias fall within the confidence bounds. Hence, it can be seen that there is no statistically significant impact of vias on the failure rate for this dataset.

2.5 Determination of the Relationship Between Lifetime and Probability Points

The traditional approach to determine lifetime for a structure (after area scaling and projection to use conditions) for a fixed low probability point, such as 0.0001, involves analyzing lifetime at a single value of distance between the lines. For example, the probability of 0.0001 corresponds to -9.21 on the Weibull scale. Simply, the lifetime that corresponds to this probability point is determined, as illustrated in Figure 2.7(c).

Die-to-die linewidth variation introduces a second dimension into the problem, since these two simultaneous mechanisms degrade lifetime. Specifically, when considering variation in linewidth, the characteristic lifetime varies in accordance with

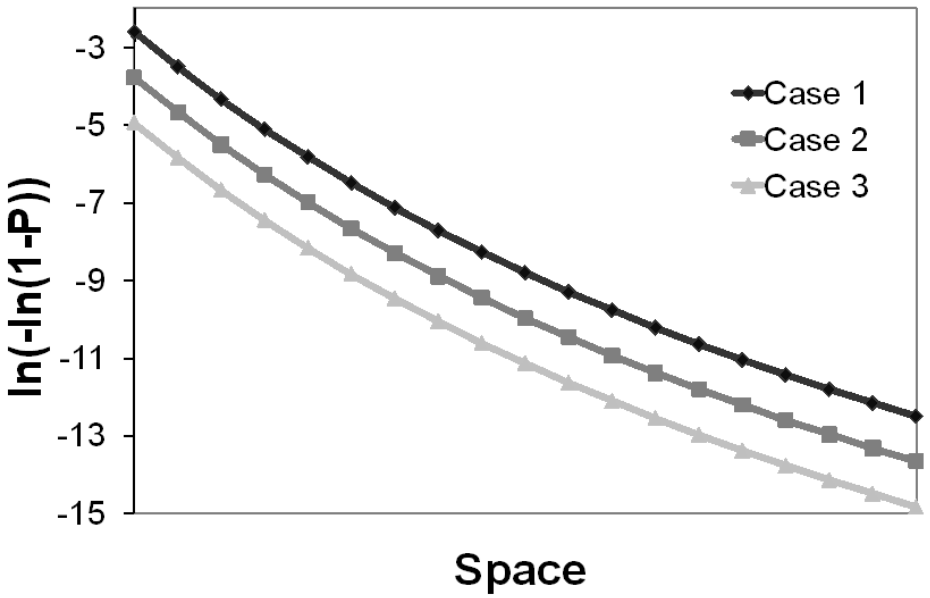


Figure 2.11: Probabilities of having a lifetime not as bad as a fixed value (labeled as Case 1 to 3) as a function of distance between the lines for the $\sqrt{E}$ model, where the fixed lifetime requirement is longest for Case 1 and shortest for Case 3.

Equations (2.10) and (2.11). Figure 2.11 illustrates probability versus distance between the lines for several fixed values of lifetime.

It can be seen that the probability of having a lifetime worse than a fixed value increases drastically as the distance between the lines decreases.

When the distance between lines varies, the probability of having a lifetime worse than a specified value involves integrating the probabilities in Figure 2.11 over the distribution of values of distance between the lines.

Thus, to account for die-to-die linewidth variation, first the relationship between lifetime and probability for specific values of distance between the lines is found using the traditional method illustrated in Figure 2.7(c). Then, for a fixed lifetime target, the probabilities are integrated over the probability density function of the distance between the lines, to find the probability associated with each fixed lifetime target.

For a normal distribution of distance between the lines, this integral is a function

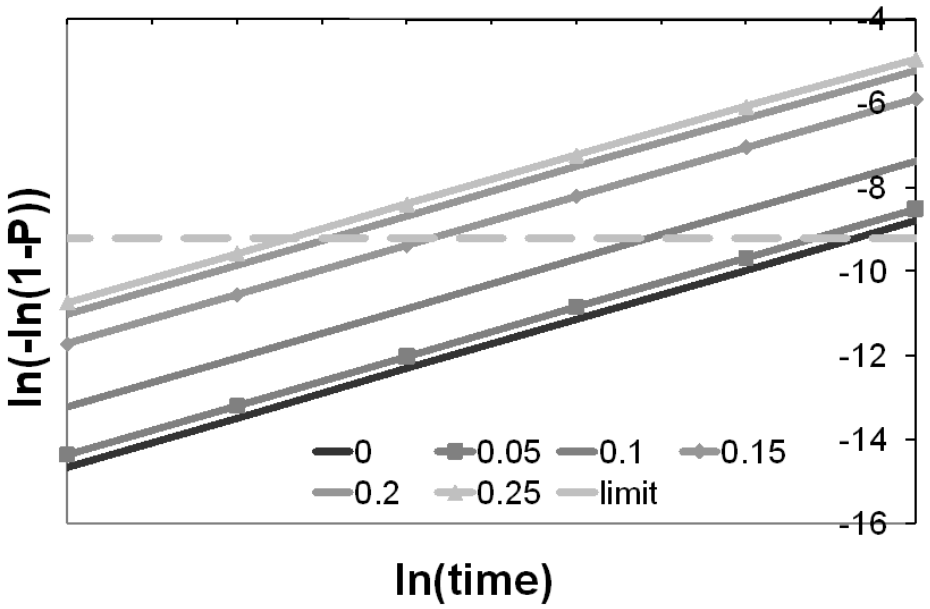


Figure 2.12: Lifetime as a function of the probability of failure for the $\sqrt{E}$ model, and as a function of the ratio of the standard deviation of variation to the nominal linewidth.

of the standard deviation of the variation. The impact of variation on lifetime is shown in Figure 2.12.

It can be seen that when variation approaches 10%, orders of magnitude improved lifetimes are required to achieve the same probability of failure. Nevertheless, the lifetime predicted at the 0.0001 probability point using this methodology, with either the E or the $\sqrt{E}$ model, is about an order of magnitude better than predicted with the traditional method, which does not account for die-to-die linewidth variation.

In addition, the results are almost insensitive to the choice of model, the E model or the $\sqrt{E}$ model, at test conditions; however, the choice of model does impact the projection to use conditions. Note that the expected lifetime depends on two parameters: print bias and random die-to-die variation in distance. Extracted die-to-die variation in distance between the lines was found to be insensitive to bias. However, lifetime projections at small percentiles are sensitive to print bias, as illustrated in

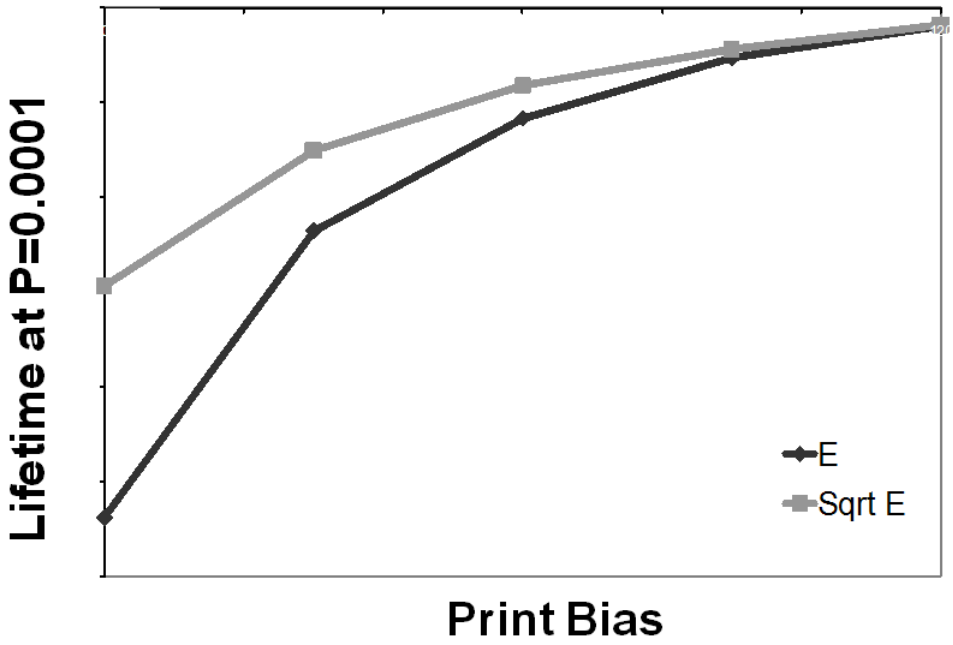


Figure 2.13: Expected lifetime at the 0.0001 probability point, as a function of print bias, for the fixed (optimum) value of die-to-die distance variation.

Figure 2.13. Hence, it is important to verify print bias with additional data prior to making lifetime projections.

2.6 Conclusion

Test structures were designed and implemented to detect the impact of interconnect geometries on backend low-k dielectric breakdown. In this chapter, the impact of field enhancement at the tips of combs, in wrap around structures, and at vias, was analyzed. In order to address these problems, an analysis methodology was proposed to separate the impact of area and field enhancement at the tips of the combs, and it was shown that there is no significant impact of field enhancement at the tips, due to wrap around structures, and at vias.

This chapter also showed that failure rate data can be used to not only extract Weibull parameters, η and β, but also the standard deviation of die-to-die linewidth variation. The proposed methodology involves, first, extracting β using area scaling and the characteristic lifetime, η, as a function of the area ratio. Then, the slope at the characteristic lifetime is matched with the data to find die-to-die linewidth variation, which degrades the slope of the Weibull curves in measured data. The extraction of die-to-die linewidth variation takes advantage of test structures which vary the distance between the lines, to determine the relationship between distance and the characteristic lifetime.

Finally, it was noted that the determination if a process satisfies a lifetime requirement should take into account both the standard Weibull statistics that characterize the failure rate of a population, η and β, and also die-to-die linewidth variation. It was shown that as variation becomes large, the lifetimes that achieve the same target probability of failure are orders of magnitude lower than the case without die-to-die linewidth variation.

CHAPTER III

IMPACT OF LINEWIDTH ON LOW-K TDDB IN COPPER INTERCONNECTS

3.1 Introduction

This chapter looks at the variation in backend low-k dielectric breakdown times as a function of metal width, a parameter that is not supposed to impact failure rates. Metal width is the width of the Cu interconnect lines drawn on the mask. Any change in the actual linewidth on chip, from the linewidth drawn on the mask, will also cause a change in actual linespace between the interconnects. Note that the terms metal width and linewidth are used interchangeably. Similarly, the terms space and linespace are used interchangeably. The ratio of the area of interconnect lines on the mask to the area of the mask is referred to as pattern density. The test structures for this chapter use a timed etch rather than an etch stop layer. A cross-section of the test structure is shown in Figure 3.1.

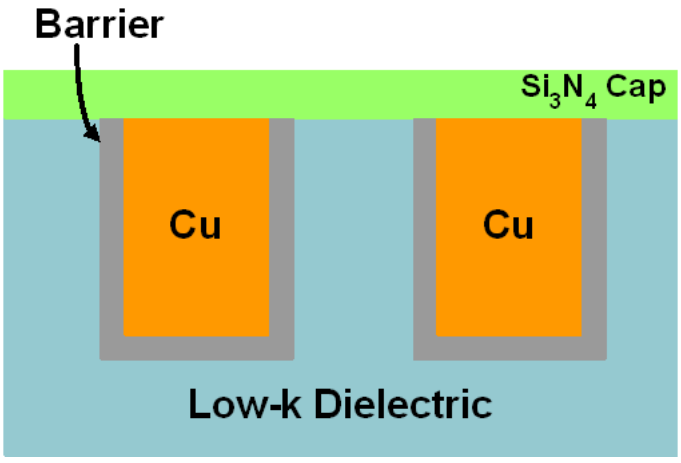


Figure 3.1: Cross-section of the copper/low-k interconnect test structure used for stress tests.

As with the previous chapter, the analysis in this chapter is built around the Weibull distribution. However, in this chapter the plots of lifetime distributions have been drawn between ln(Time) on the x-axis, as in the previous chapter, and %Failed on the y-axis, a departure from the previous chapter. Note that the 0 on the $\ln(-\ln(1-P))$ scale equals 63.2% on the %Failed scale. Thus the characteristic lifetime, η, of the Weibull distribution is the time at which 63.2% of the test structures have failed. An example is shown in Figure 3.2.

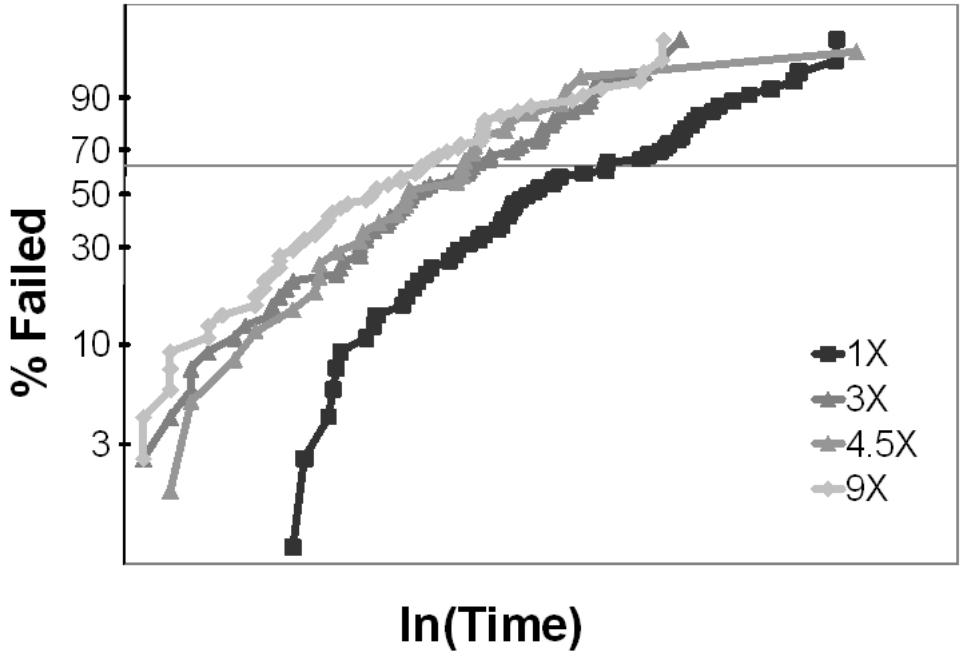


Figure 3.2: Example Weibull plot with a modified y-axis for comb test structures with four areas: $1X$, $3X$, $4.5X$, and $9X$.

This chapter begins with a summary of prior work on modeling variation in failure rates as a function of linewidth in the following section. Section 3.3 describes the test structures that were used in this work. Section 3.4 summarizes the TDDB measurement results. Section 3.5 considers several possible explanations of the observed variation in characteristic lifetime as a function of linewidth. Section 3.6 uses the results in Section 3.4 to create a model of characteristic lifetime as a function of

linewidth. Section 3.7 concludes the chapter with a summary.

3.2 Prior Work and Motivation

The two dominant models of TDDB lifetime, the E Model [42] and the $\sqrt{E}$ Model [49, 50], relate time-to-failure to electric field. In both models, the factors that determines TF for structures manufactured using the same low-k dielectric are the electric field (E) and the temperature (T). For a fixed potential, V, electric field in back-end structures is a function of the distance between the interconnect lines, termed linespace (S),

$$E = \frac{V}{S}. \tag{3.1}$$

In the used test structures and in prior work on this topic, S is constant, and only the linewidth is varied. In prior work, with $180nm$ technology, experimental data indicated that time-to-failure was a function of linewidth [63]. Analysis found that the difference in time-to-failure was due to a physical difference in the distance. The data was analyzed to determine an explanation for the difference in distance. The explanation that best matched the data was microloading in etch [63]. The microloading effect was explained as a sensitivity of etch rate to pattern density [120, 121]. However, the test structures used to analyze the impact of metal linewidth confounded the impact of linewidth with pattern density, as can be seen in Figure 3.3. Figure 3.3 shows that whenever linewidth is increased, while keeping linespace constant, the pattern density also increases.

Specifically, let W_i be the linewidth of the i^{th} test structure, and the distance between the adjacent lines be S. The pitch, P_i, for the i^{th} test structure is

$$P_i = W_i + S. \tag{3.2}$$

The pattern density, ρ_i, for the i^{th} test structure, is defined as the ratio of the total metal area in the test structure to the total area of the test structure.

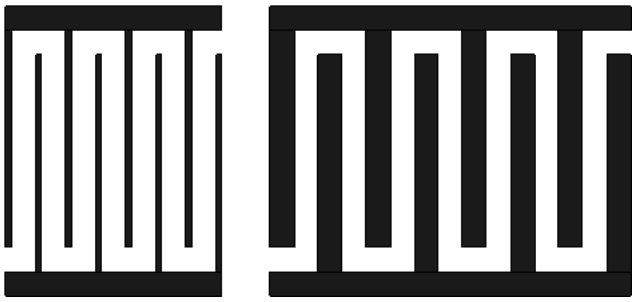


Figure 3.3: Test structures that vary both linewidth and density concurrently (top view). Both the test structures have the same linespace.

Let L be the length of the lines. If there are n lines in each comb, then the total number of lines in the structure is $N = 2n$. Let's denote the metal area as A_{Mi} for the i^{th} test structure. Then

$$A_{Mi} = NW_iL. \tag{3.3}$$

The total area of the dielectric, A_D, is

$$A_D = (N-1)SL. \tag{3.4}$$

The total area of the i^{th} test structure is

$$A_i = A_{Mi} + A_D = L(NW_i + (N-1)S). \tag{3.5}$$

From (3.3) and (3.5)

$$\rho_i = \frac{A_{Mi}}{A_i} = \frac{W_i}{W_i + \frac{N-1}{N}S} \ . \tag{3.6}$$

Now, let's suppose that one test structure has $W_i = W$. Then, if another test structure has $W_i = QW$, its pattern density is

$$\rho_Q = \frac{W}{W + \frac{N-1}{NQ}S} \ . \tag{3.7}$$

In the limit, as the number of lines, N, becomes large,

$$\rho_Q = \frac{1}{1 + \frac{S}{WQ}} \ . \tag{3.8}$$

Clearly, as Q increases, ρ_Q also increases.

Hence, although the theory associates the time-to-failure difference with pattern density, it was not conclusively verified that pattern density, rather than linewidth, produced the time-to-failure difference. The purpose of this chapter is to distinguish between these two factors.

3.3 Test Structure Design

Test structures that vary metal linewidth and density separately were designed with the aim to distinguish the impact of linewidth and density. Two of the test structures that vary linewidth while keeping the linespace constant, and hence vary the pattern densities, are shown in Figure 3.3. Two other test structures have the same density and are shown in Figure 3.4. One has non-uniform linewidth, with thin and wide lines that match those in Figure 3.3.

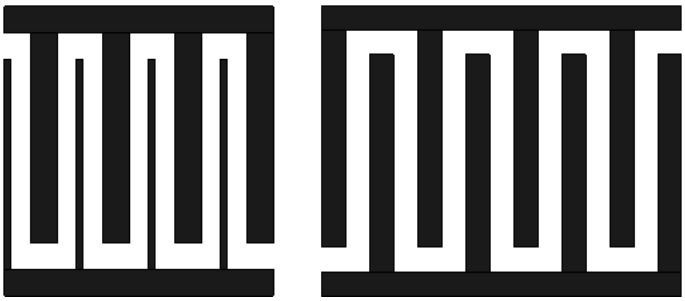


Figure 3.4: A test structure pair that can distinguish between the impact of density and linewidth (top view). Both the test structures have different linewidths but the same pattern density.

The following terminology is used. The structure with minimum linewidth is referred as $1X$, since the drawn width of the lines is $1X$. The structure with linewidths that are N times the minimum linewidth is referred as NX. $3X$ and $5X$ test structures are used. The test structure with non-uniform linewidth is referred as $1X/5X$, since one of the combs has $1X$ linewidth and other has $5X$ linewidth. Note that $1X$, $3X$ and $5X$ test structures have different linewidths, but the same linespace on the mask. Test structures $3X$ and $1X/5X$ have the same pattern density.

If TF is a function of density, then TF should be the same for the two test structures in Figure 3.4. On the other hand, if linewidth is the cause of the TF difference, then there will be a difference in TF distributions for the two structures. Furthermore, TF of the non-uniform structure in Figure 3.4 should be predictable using TF distributions of the two test structures in Figure 3.3 that match the linewidths of the fingers of the non-uniform structure, using area scaling, combined with the method proposed, in the previous chapter, to predict the failure rate when there are two independent failure mechanisms [122, 123].

3.4 Test Results

3.4.1 Uniform Linewidth Test Structures

Figure 3.5 shows the failure rate distributions for the test structures with $1X$, $3X$, and $5X$ linewidths. These test structures vary both linewidth and density. Note that for the test structures in Figure 3.5, only the linewidth and the pattern density are varied and the distance between the lines has remained constant.

Figure 3.5 also shows a model and confidence bounds for the $1X$ linewidth dataset, indicating that the changes in the characteristic lifetime cannot be attributed to random variation. The model was computed based on data from $1X$, $3X$, $4.5X$, and $9X$ area test structures [122]. The shape parameter, β, cannot be extracted directly, since the Weibull curves are impacted by die-to-die linewidth variation. The die-to-die linewidth variation was extracted by accounting for the difference between β (extracted by area scaling) and the slope of the Weibull curves [122, 123]. Figure 3.5 shows that the improvement in characteristic lifetime for wide lines is statistically significant.

β was also extracted for each of the test structures. This was done by curve fitting, assuming the same die-to-die variation computed for the $1X$, $3X$, $4.5X$, and $9X$ area test structures in [122, 123]. β decreases for the test structures with wider lines.

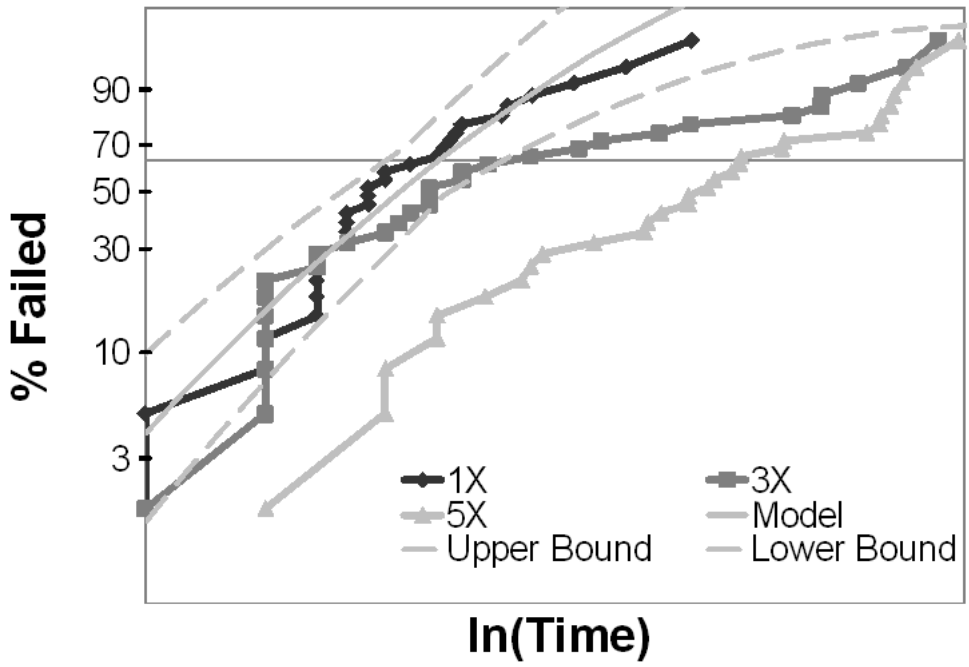


Figure 3.5: Time-to-failure distributions for test structures with $1X$, $3X$, and $5X$ linewidths. 90% confidence bounds are added for the $1X$ test structure.

3.4.2 Non-Uniform Linewidth Test Structures

Figure 3.6 compares data from the non-uniform $1X/5X$ structure with the $3X$ structure, which matches its pattern density. Their failure distributions do not match. Consequently, pattern density does not appear to be a major factor causing a difference in lifetime.

$1X/5X$ test structure is a combination of $1X$ and $5X$ linewidth test structure. It was checked if the data from the $1X$ and $5X$ models can be combined to predict the results for the non-uniform $1X/5X$ structure.

In order to generate a Weibull plot for the combination of the two structures, a defect generation function, $\lambda(t)$, must be computed. The cumulative Weibull distribution

$$P(TF) = 1 - \exp\left(-\left(\frac{TF}{\eta}\right)^{\beta}\right), \tag{3.9}$$

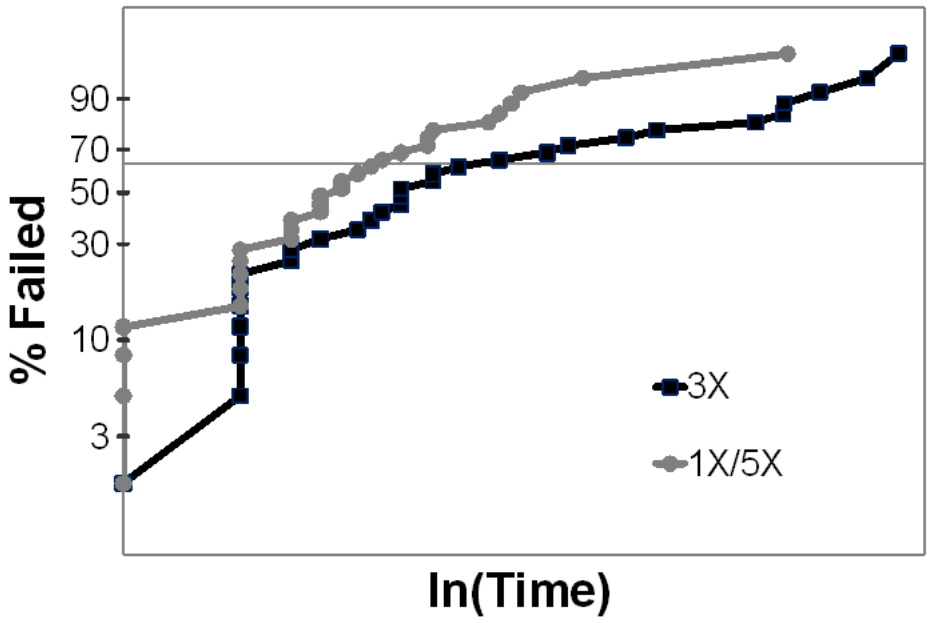


Figure 3.6: Failure rate distribution comparison for test structures with equal density.

where P is probability, and the cumulative Poisson model

$$P(TF) = 1 - \exp(-\lambda(TF)A), \tag{3.10}$$

where A is the test structure area, are used to obtain $\lambda(TF)$.

The defect generation functions for the $1X$ and $5X$ structures are

$$\lambda_{1X}(TF) = \frac{1}{A}\left(\frac{TF}{\eta_{1X}}\right)^{\beta_{1X}}, \tag{3.11}$$

and

$$\lambda_{5X}(TF) = \frac{1}{A}\left(\frac{TF}{\eta_{5X}}\right)^{\beta_{5X}}. \tag{3.12}$$

The total number of defects at any time is $\lambda(TF)A$. Therefore, the number of defects for the combined structure with $A/2$ of $1X$ and $A/2$ of $5X$ is

$$\lambda(TF)A = \lambda_{1X}(TF)\frac{A}{2} + \lambda_{5X}(TF)\frac{A}{2}, \tag{3.13}$$

and the cumulative probability density function is

$$P(TF) = 1 - \exp\left(-\frac{1}{2}\left(\left(\frac{TF}{\eta_{1X}}\right)^{\beta_{1X}} + \left(\frac{TF}{\eta_{5X}}\right)^{\beta_{5X}}\right)\right). \tag{3.14}$$

This equation is converted to Weibull plot format, as follows:

$$\ln(-\ln(1-P(TF))) = \ln\left(\frac{1}{2}\left(\left(\frac{TF}{\eta_{1X}}\right)^{\beta_{1X}} + \left(\frac{TF}{\eta_{5X}}\right)^{\beta_{5X}}\right)\right). \tag{3.15}$$

Note that Weibull plots for 1X and 5X structures are used and for these plots

$$y_{1X} = \ln(-\ln(1-P_{1X}(TF))) = \ln\left(\left(\frac{TF}{\eta_{1X}}\right)^{\beta_{1X}}\right), \tag{3.16}$$

and

$$y_{5X} = \ln(-\ln(1-P_{5X}(TF))) = \ln\left(\left(\frac{TF}{\eta_{5X}}\right)^{\beta_{5X}}\right), \tag{3.17}$$

where y_{1X} and y_{5X} are y-axis values from the Weibull plots for the $1X$ and $5X$ structures, respectively. Hence,

$$\ln(-\ln(1-P(TF))) = \ln\left(\frac{\exp\left(y_{1X}(TF)\right)}{2} + \frac{\exp\left(y_{5X}(TF)\right)}{2}\right). \tag{3.18}$$

Figure 3.7 is computed by interpolation at each point, t, in the Weibull plots for the $1X$ and $5X$ structures using Equation (3.18). As can be seen from Figure 3.7, the model and the data do not match.

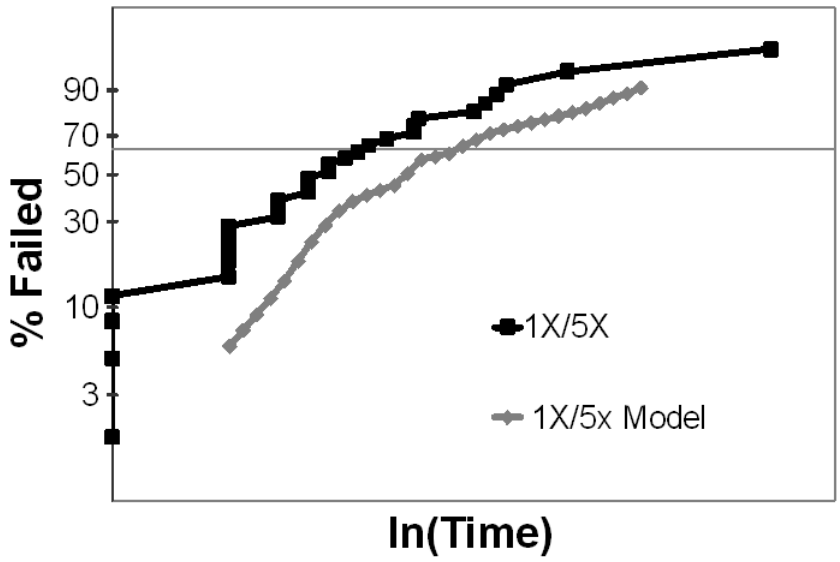


Figure 3.7: Predicted failure rate for the $1X/5X$ structure, based on data from the $1X$ and $5X$ structures.

3.4.3 Linewidth and Pattern Density

Figure 3.8 shows the plot of the characteristic lifetime vs. linewidth. Linewidth has a clear impact on the characteristic lifetime, and the characteristic lifetimes increase with linewidth. However, when linewidth is varied, while keeping the line spacing constant, the pattern density also changes. Figure 3.9 plots the characteristic lifetime vs. pattern density. Both Figure 3.8 and Figure 3.9 provide a correlation between time-to-failure and geometry.

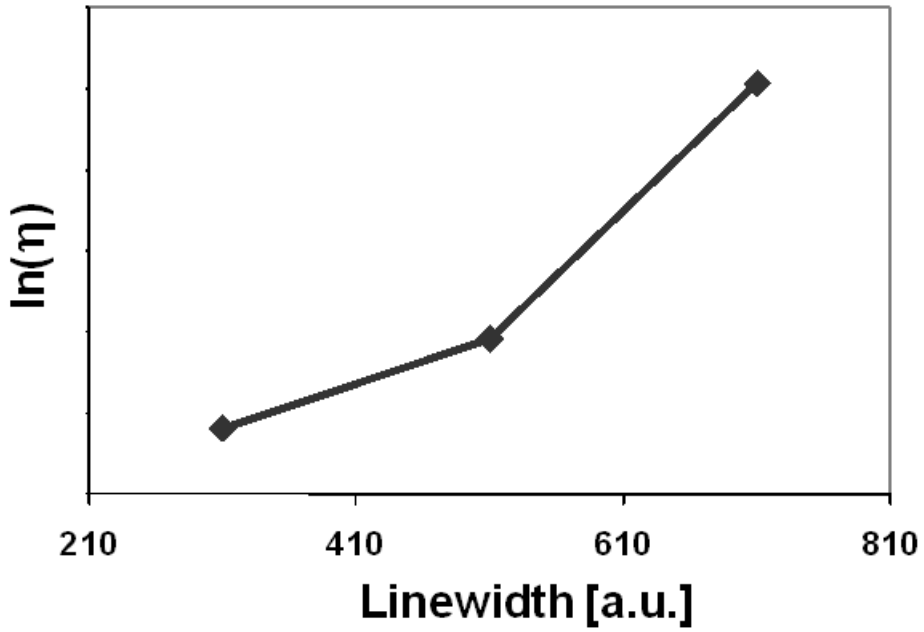


Figure 3.8: Characteristic lifetime as a function of linewidth.

3.5 Analysis of Potential Causes of Variation

This section considers some possible explanations for the variation in characteristic lifetime as a function of either linewidth or pattern density.

3.5.1 Variation as a Function of Printed Geometry

Manufactured geometries were collected for the used test structures using scanning electron microscopy (SEM), shown in Figure 3.10. The data extracted is shown in

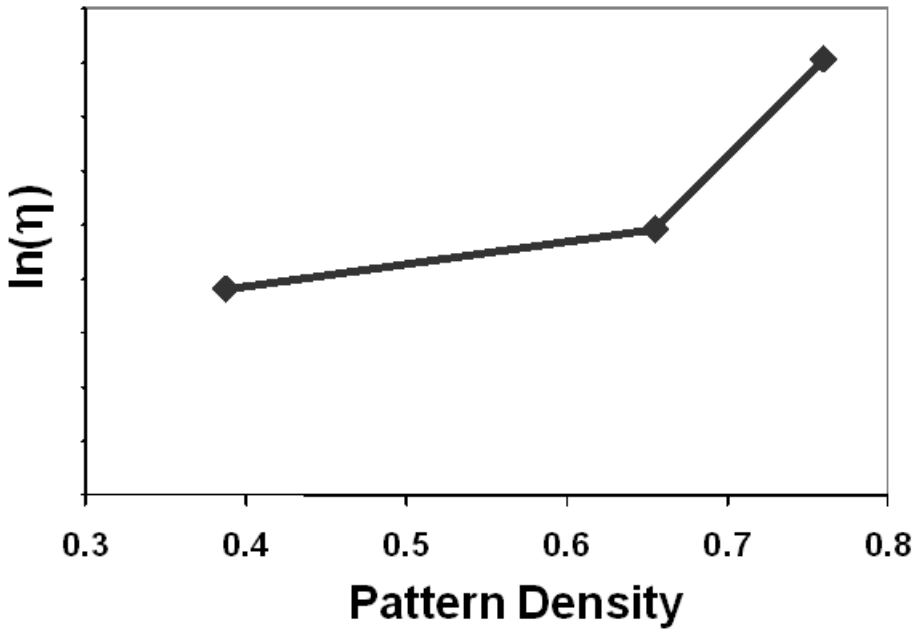


Figure 3.9: Characteristic lifetime as a function of pattern density.

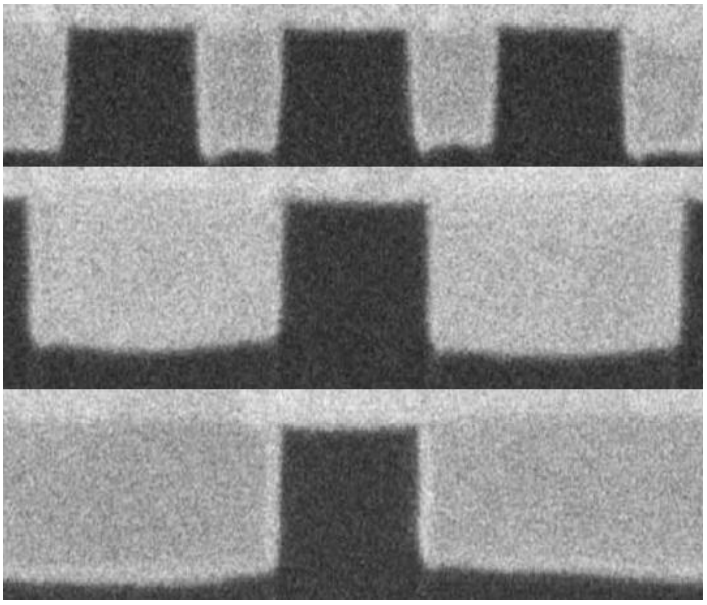


Figure 3.10: SEM images for the uniform linewidth test structure. Cross section of $1X$ linewidth test structure is at the top. The specified linewidth increases from the top cross section to the bottom. Note that the pictures are cut-outs from the uniform linewidth test structures. When computing dimensions, to reduce errors, the pitch is assumed to be known.

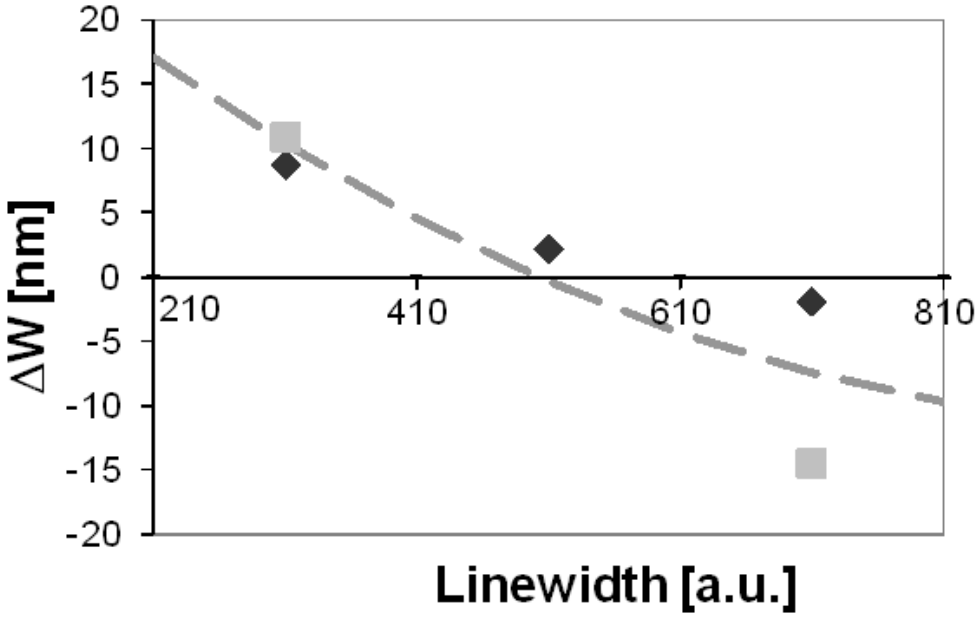


Figure 3.11: The manufactured shift in linewidth as a function of linewidth on the mask. The light grey dots correspond to the non-uniform test structure and the black dots correspond to the uniform linewidth test structures. The model is computed with regression.

Figure 3.11.

In this graph, the linewidth difference is defined as

$$\Delta W = W_A - W_D. \tag{3.19}$$

W_A is the actual linewidth and W_D is the drawn linewidth. In Figure 3.11 grey dots correspond to the non-uniform linewidth test structure and give the ΔW for the $1X$ comb and the $5X$ comb in the $1X/5X$ test structure. The black dots correspond to the uniform test structure with $1X$, $3X$, and $5X$ linewidths. The graph indicates that the narrow lines are wider than drawn, and the wide lines are narrower than drawn. Potential explanations for variation in the printed linewidth, lithography, etch, and field enhancement, are investigated next.

3.5.2 Lithography

The aerial image of a test structure varies with pattern density because of the optical proximity effect. The optical proximity effect varies as a function of focal depth and pitch. The radius of influence is around $400nm$ for an illumination system with a wavelength of $193nm$. Hence, the optical proximity effect can influence the narrow lines, but is less likely to influence the wider ones. It tends to increase the linewidths of dense structures, depending on exposure dose. The narrow linewidths are the least dense structures, and therefore they should be the most narrow. The manufactured linewidth difference should be the most negative. The trend in Figure 3.11 is the opposite and is inconsistent with the optical proximity effect, where the lines with larger linewidths have smaller values of ΔW, which translates to larger distances between the lines.

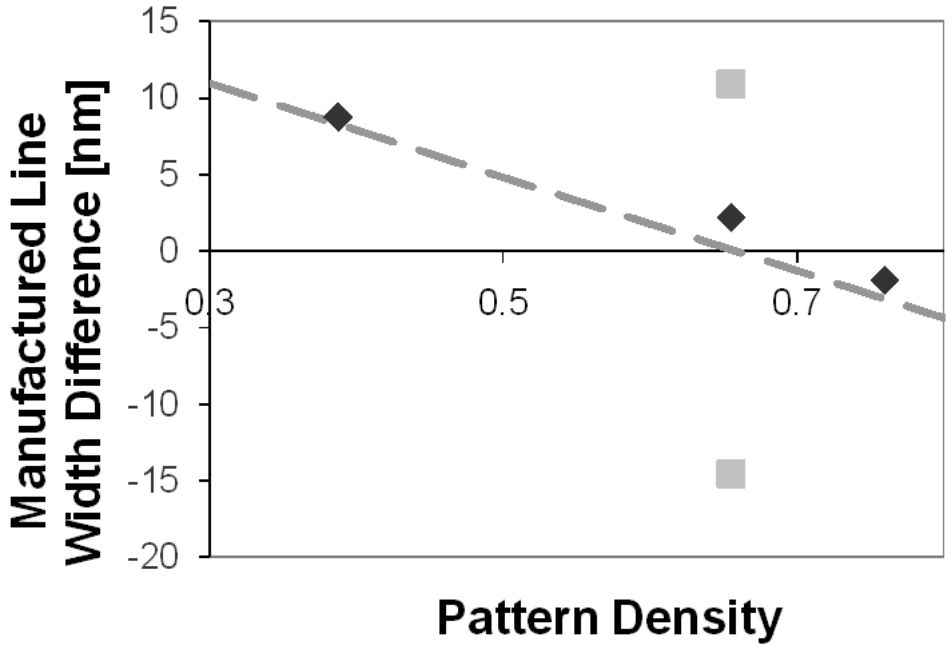


Figure 3.12: The variation in linewidth in the fabricated chip as a function of pattern density. The black dots correspond to the uniform test structures and the grey dots correspond to the non-uniform test structure.

Flare, the deterioration of printed image because of light scattering, is another source of linewidth variation. Flare increases with decreasing feature size in the sub-$100nm$ regime, with smaller linespaces showing as much as 50% narrowing [124]. Flare is a function of pattern density. Figure 3.12 is a plot of the linewidth difference vs. density and does not indicate a consistent relationship between pattern density and difference in linewidth. In fact, the test structures showed a trend opposite to the one previously observed for flare. Hence, flare is unlikely to be an explanation for the variation in linewidth.

3.5.3 Etching

Etch rate and etch selectivity have been shown to be strongly dependent on pattern density. Previous models have established a link between the linespace, linewidth, etch rate, and mean-time-to-failure (MTTF) [63]. These models use Mogab's model for etch rate [120]. Mogab's model predicts a decrease in the etch rate, ER, and consequently an increase in the linespace, as the linewidth increases, i.e.,

$$ER \propto \frac{1}{A_E},$$

and

$$ER \propto \frac{1}{W}, \tag{3.20}$$

where A_E is the etchable area and W is the linewidth.

However, unlike in [63], the collected dataset shows variation in the height of the structures. This is summarized in Figure 3.13. This is because the process uses a timed etch, rather than an etch stop layer. The data in Figure 3.13 shows a correlation between linewidth and line height. Based on this data the line height was assumed proportional to the etch rate. A model was computed for etch rate, and is shown in Figure 3.14.

Two factors can create pattern dependencies in etch [125]: microloading, which depends on pattern density, and aspect ratio-dependent reactive ion etch lag.

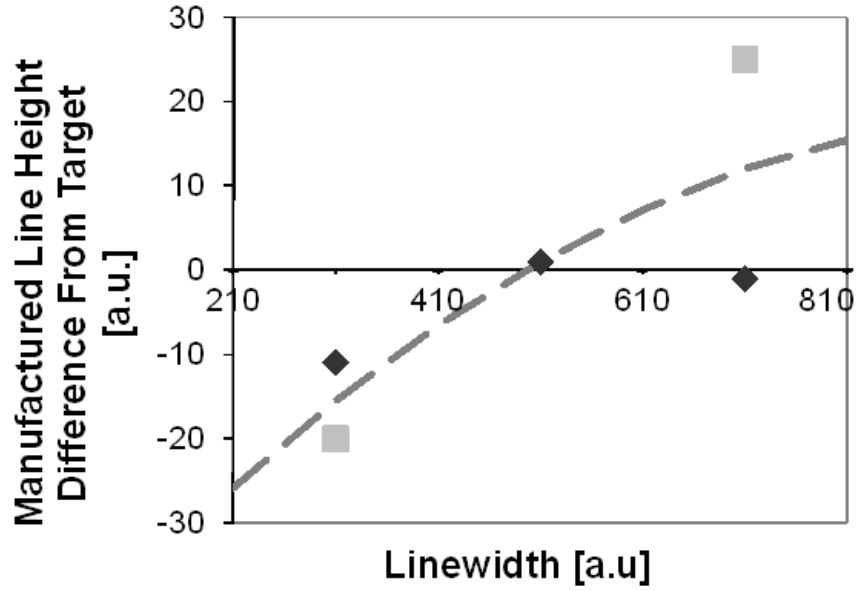


Figure 3.13: The manufactured line height as a function of linewidth. The black and grey dots correspond to the uniform and non-uniform test structures, respectively. The model is computed with regression.

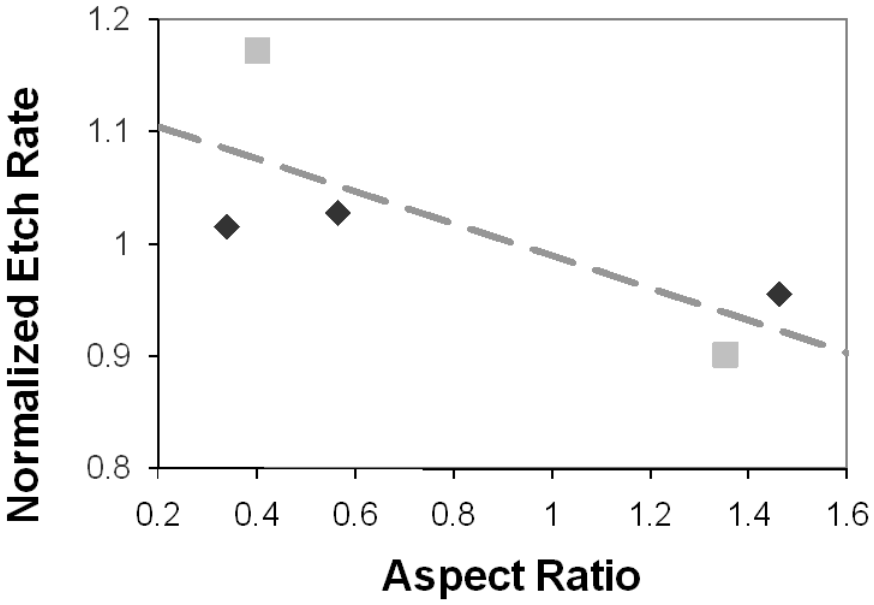


Figure 3.14: The etch rates for test structures were found to vary as a function of aspect ratio. The aspect ratio is computed using measured data. The black dots and grey dots correspond to the uniform and non-uniform structures, respectively. The model is computed with regression.

Pattern density causes spatial variation in etch rate by changing the concentration of reactants in areas with different pattern densities, as different features compete for reactants over short distances [125]. Taylor et al. [126] and Abrokwah et al. [125] report a decrease in etch rate with increasing pattern density. The test structures showed an increase in etch rate with increasing pattern density, Figure 3.15, as in [127,128] if there is any relationship at all, opposite to the trend reported in [125,126].

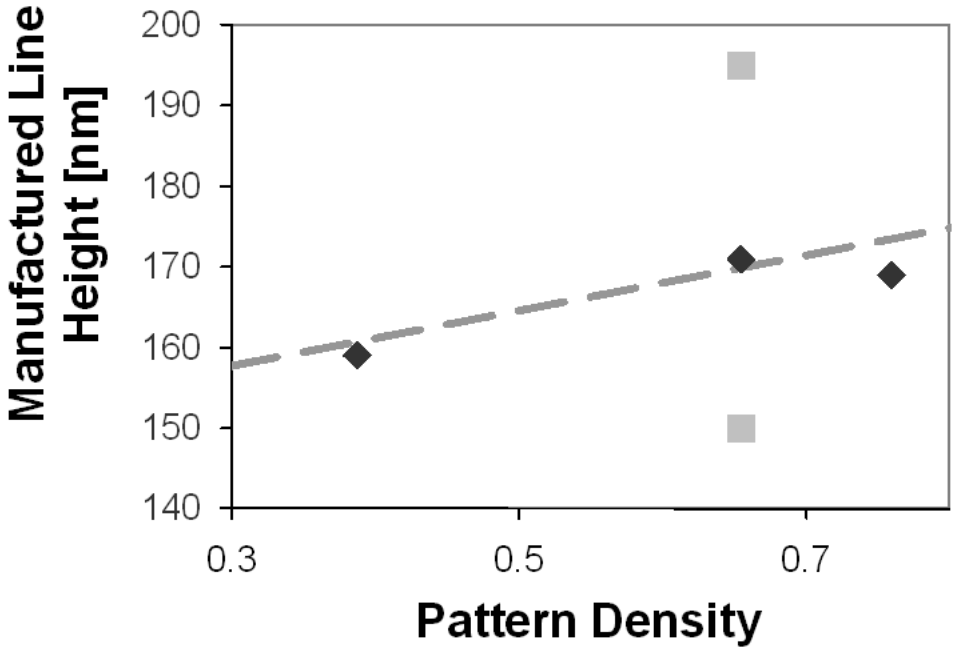


Figure 3.15: The manufactured line height as a function of pattern density.

Aspect ratio dependant etching (ARDE) manifests itself in submicron feature sizes having high aspect ratios (feature height/feature width). In the presence of ARDE, higher aspect ratio trenches etch slower [129]. Figure 3.14 shows the etch rates for the test structures, along with their aspect ratios.

When the etch rate increases with trench size, this indicates that the process is chemically-controlled. Ion bombardment is not controlling the etch, but rather the concentration of etchant species entering the trench increases with increasing trench

width. Therefore, as the trench width increases, more etchant can enter the trench (since etchant arrives at random angles), thereby increasing the etch rate [125, 129].

The data indicates that the etch rate is composed of two different etch rate components, the lateral etch rate and the vertical etch rate, both of which depend on aspect ratio.

The trend observed in the actual linewidths can be attributed to the lateral component of etch rate. If the line heights are taken as an indicator of vertical etch rate, then the line heights indicate that the vertical etch rate decreases with increasing aspect ratio, while the lateral etch rate increases.

Non-parallel incident ion trajectories can cause the sidewalls to taper. The smaller linewidth test structures showed larger taper. However, non-parallel incident ions often give rise to barreling which wasn't observed in the used test structures [130]. It is possible that the test structures with wider lines may be protected by polymer deposition on the sidewalls. The narrow lines may be less protected by polymer deposition, and therefore may experience more lateral etch.

In particular, initially, as more etchant enters the trench, the lateral etch rate also increases. Both more Fluorine (F) and polymer, containing Carbon (C), enter the trench. The polymer is deposited on the sidewall and at the bottom of the trench [131]. This polymer deposition increases as linewidth increases. At the bottom of the trench, ion bombardment removes the deposit, making more etchant (F) available, increasing the etch rate and depth [131]. Since the sidewalls do not receive much, if any, ion bombardment, the sidewall deposits build up and become thicker with increasing trench width. The sidewall deposition protects the sidewall from additional etching. The polymer deposits are thinner on the narrower trenches. Hence, the more narrow trenches are less protected by polymer deposition, and consequently suffer from increased lateral etching. This increase in lateral etching reduces the linespace for the narrow lines, by increasing their linewidth, leading to observed TF trends.

The impact of the lateral component of the etch rate on linewidth as a function of aspect ratio is shown in Figure 3.16.

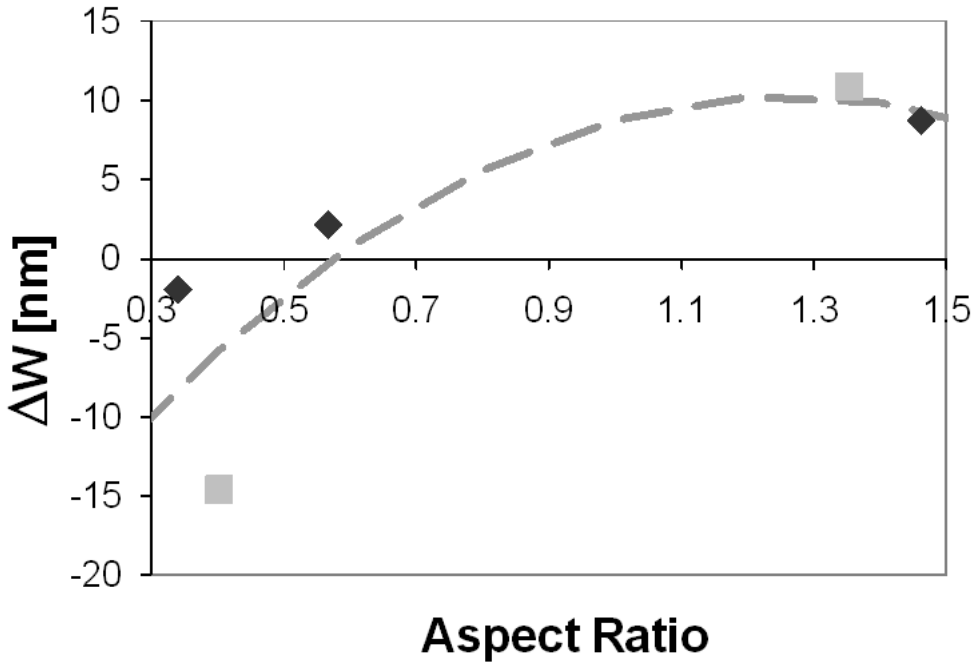


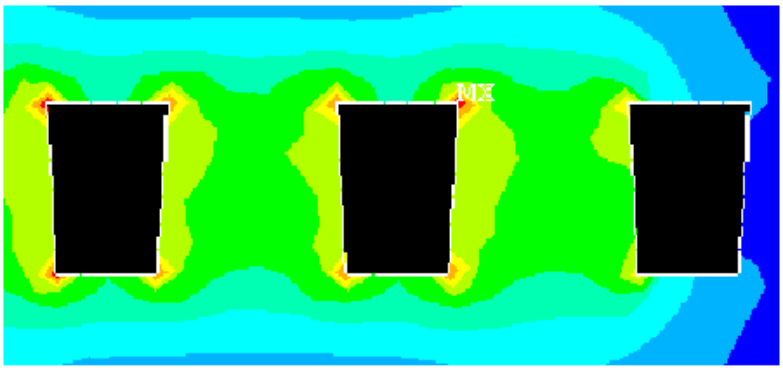
Figure 3.16: The variation in linewidth in the fabricated chip as a function of aspect ratio. The black dots and lines correspond to the uniform test structures, and the grey dots correspond to the non-uniform test structure. The model is fit by regression.

Lateral etching is the main factor that affects the characteristic lifetime. Increased lateral etching in the narrow trenches leads to shorter TF because of the reduced linespace.

3.5.4 Variation as a Function of Electric Field Enhancement

A potential cause of variation in characteristic lifetime is electric field enhancement due to fringing effects.

Finite element (FEM) simulations were carried out using ANSYS to determine the effect of geometry on electric field. Finite element simulations showed high electric field intensities at bends and tips. This is consistent with results in the literature [57]. Figure 3.17 shows results of finite element simulations for the uniform linewidth test



(a)

(b)

Figure 3.17: Results of finite element simulations for the (a) uniform linewidth test structure (b) non-uniform linewidth test structure. Different colors show different electric field intensities, with the darkest being the highest. Black color shows the Cu lines.

structure and the non-uniform linewidth test structure.

The locations of high electric field in finite element simulations coincide with vulnerable locations in the Cu/low-k interconnect structure. Specifically, after formation of the trench, barrier metals are blanket deposited, followed by the deposition of a Cu seed layer and Cu deposition via electroplating. After Cu deposition via electroplating, CMP is carried out to remove the excess Cu covering the dielectric. The difference in hardness between the barrier layer, the soft Cu layer, and the softer low-k dielectric layer, can lead to an uneven profile along the top edge of the trench. This uneven profile, along with high electric fields at the corners of the trench, can trigger Cu diffusion and lead to breakdown. In the literature, breakdown sites have been observed around the corners of the trenches [48, 113, 132].

The electric field distribution should indicate the potential defect sites in the

dielectric. However, the exact value of the maximum electric field is determined by the corner rounding at the corners, as noted in [48]. Results from simulations show that the maximum electric field at the corners does not follow any particular trend as linewidth increases, as shown in Figure 3.18. Hence, the role of the maximum electric field at the corners is excluded in this analysis.

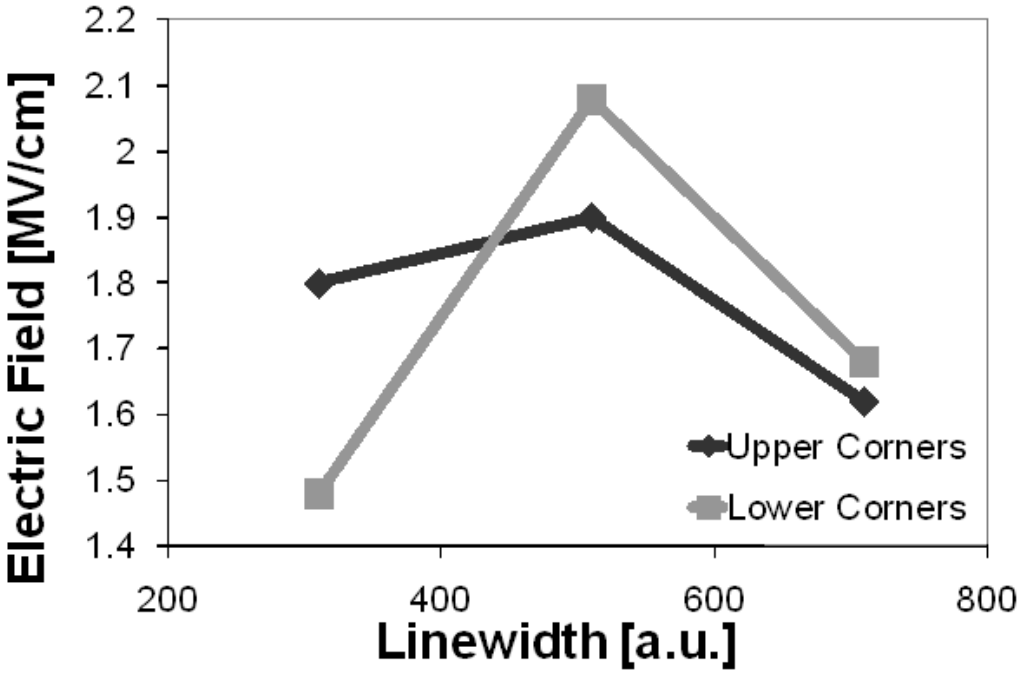


Figure 3.18: Peak electric field at the upper and lower corners of the test structures from finite element simulations.

Excluding field enhancement at the corners, the other potential sites for breakdown are the ones between the metal lines where there is sufficient uniformity in the values of the electric field intensity to cause trap generation leading to TDDB. Nanocracks or nanopores in the barrier may provide a diffusion or drift path to Cu. A compromised barrier can cause breakdown through the bulk of the dielectric. It should be noted, however, that in the used test structures no evidence of a compromised barrier was found by using Transmission Electron Microscopy (TEM), although

there are other, more sophisticated, methods of barrier characterization that can detect defective barriers [133].

Peak electric field intensities in the bulk of the dielectric (along a line centered between the interconnect lines) were computed as a function of linewidth. In these simulations, the physical dimensions of the lines were used. If the distance between the lines were to determine the maximum electric field for a fixed applied voltage, then the maximum electric field, E_M, would relate to the change in width, ΔW, as

$$E_M \propto \frac{S_D}{S_D - \Delta W} \ , \tag{3.21}$$

where S_D is the drawn linespace and $S_A = S_D - \Delta W$ is the actual linespace. This relationship is shown in Figure 3.19.

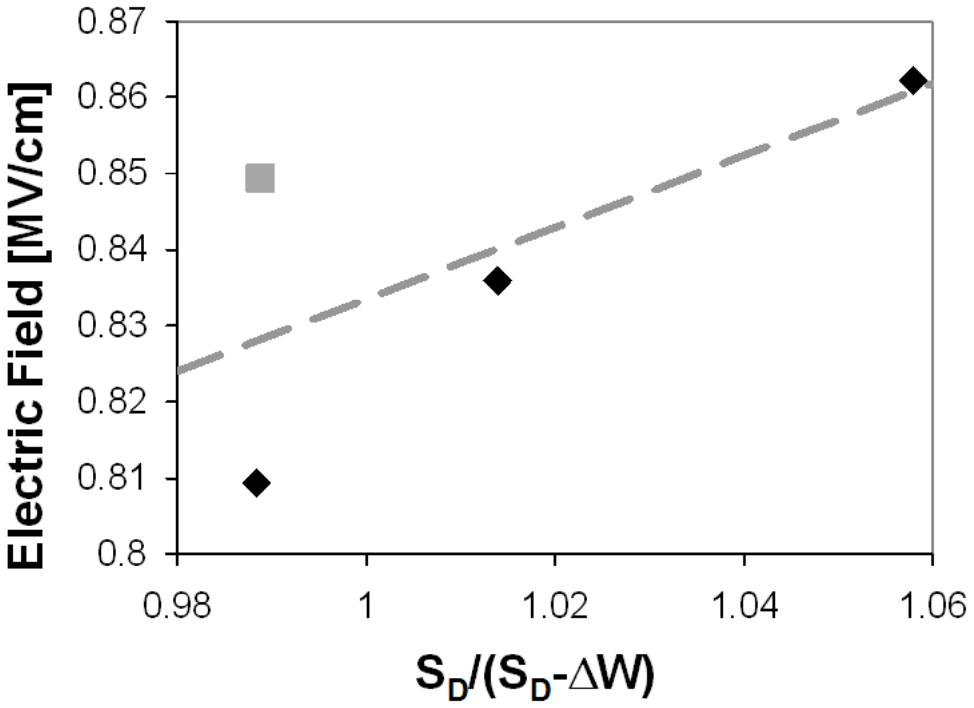


Figure 3.19: Maximum electric field at the midpoint between the lines vs. the scale factor for the shift in distance between the linespace for actual and ideal structures. The black dots and grey dots correspond to the uniform and non-uniform structures, respectively. The model is computed with regression.

Figure 3.19 shows that for the uniform linewidth test structures the maximum

electric field in FEM simulations is primarily a function of distance between the lines . It is not a function of distance for the non-uniform structure, because the wide lines are not equally spaced between the narrow lines. Therefore, the maximum electric field is a function of the minimum spacing between the lines. If the non-uniform spacing between the lines is taken into account, there is a direct relationship between electric field and (3.21), as illustrated in Figure 3.20. This shows that there is *no unexpected field enhancement.*

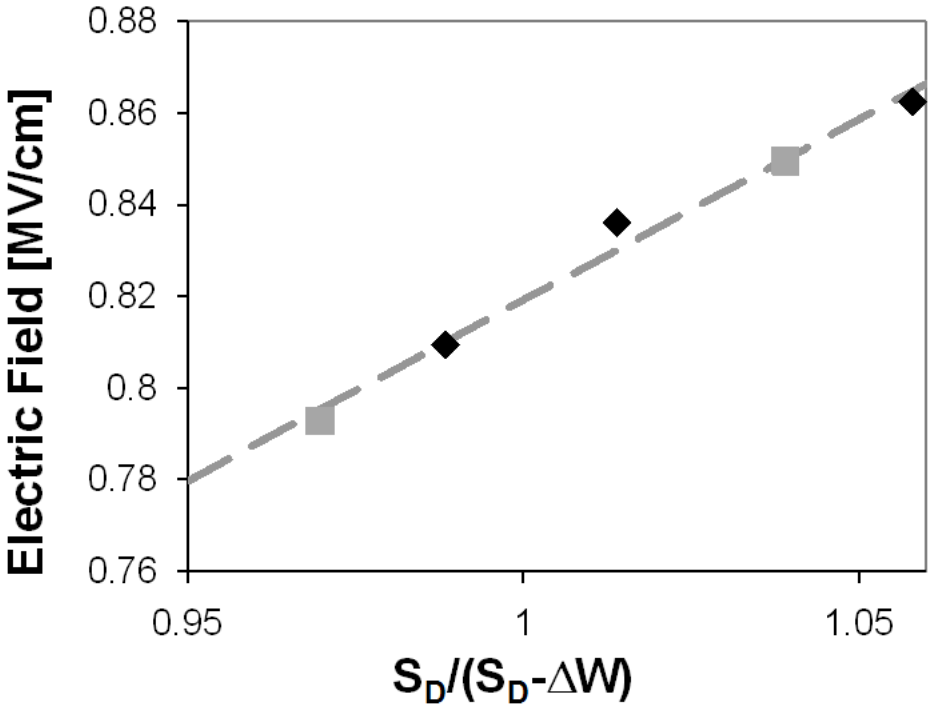


Figure 3.20: Maximum electric field at the midpoint between the lines vs. the scale factor for the shift in distance between the lines, accounting for the fact that the non-uniform structure has non-uniform linespace. The black dots and grey dots correspond to the uniform and non-uniform structures, respectively. The model is computed with regression.

3.5.5 Variation as a Function of Material Structure

Besides linewidth and electric field enhancement, there may be variation in the material structure as a function of linewidth. In this section, possible effects of stress

migration and plasma damage are considered.

3.5.5.1 Stress Migration

Linewidth has been found to relate to variation in the characteristic lifetime because of stress migration [134]. Stress migration is due to temperature variation during processing. High temperature process steps cause materials to expand. The mismatch in the temperature expansion coefficients of Cu, the barrier, and the dielectric, creates stress, which can cause problems in adhesion and potentially cracks in the interface between copper and the dielectric. Stress is a function of material volume. It has been found that larger volumes produce greater stress. Wider lines are associated with larger volumes. Hence, in [134], it was found that wider lines produce higher failure rates, due to the larger stress magnitudes in Cu. This is the opposite of the trend observed in the test data, and hence stress migration appears not to be a factor affecting out dataset.

3.5.5.2 Plasma Damage

Plasma etch and ash processes can alter surface and bulk properties of porous low-k films [135, 136]. They impact film surface chemistries, which can result in increased water absorption and surface roughness. One of the processes most likely to damage the surface is the combination of fluorocarbon etching and photoresist stripping plasmas [135]. In [135], it was found that after C_4F_8/Ar plasma treatment, the surface of a low-k material became very rough, pores opened up, and microchannels (approximately $5 - 8nm$ wide) were formed deep into the dielectric. The damage reached a depth of $165nm$. The damage was a result of bond scissioning and atomic rearrangements following bombardment by inert Argon (Ar) atoms. The dangling bonds after plasma treatments can lead to enhanced moisture absorption and increased conduction through the dielectric. Barrier metals can diffuse into the dielectric [135,137,138], reducing the effective distance between the lines. However, the experiments in [135]

demonstrated damage due to direct bombardment by plasma etch of the top surface of the dielectric. The sidewall is less likely to experience direct bombardment during etch, and consequently is less likely to be damaged to the extent shown in [135]. Hence, there is no concrete evidence that the test structures experienced plasma damage.

3.5.6 Barrier Integrity

TEM and Scanning TEM (STEM) analysis were carried out on cross sections of the tests structures to observe the integrity of the barrier metal. Figure 3.21 shows results of TEM analysis. Figure 3.22 shows results from STEM analysis. No defects were found in the barrier of the used test structures.

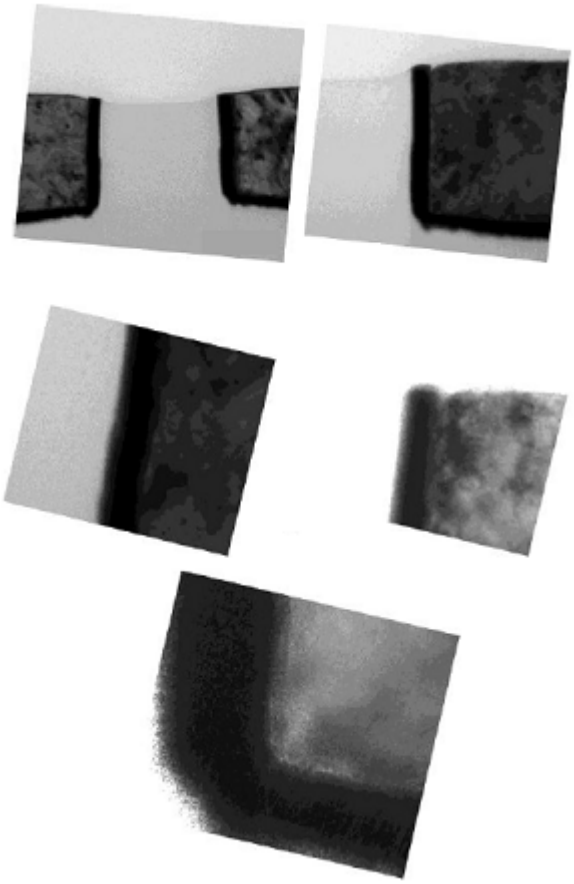


Figure 3.21: TEM micrographs of cross-sections of a test structure.

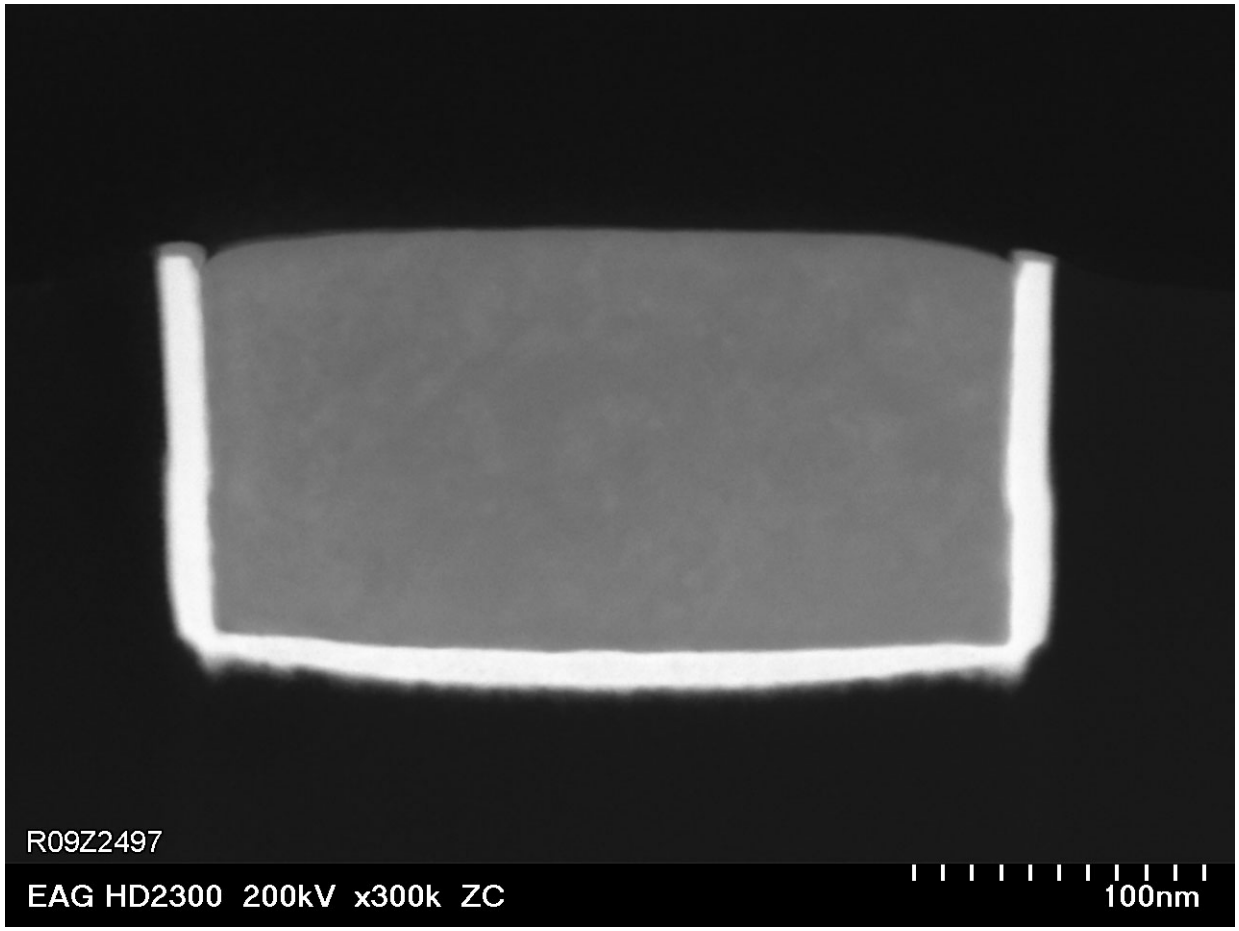


Figure 3.22: STEM micrographs of a cross-sections of $3X$ linewidth test structure.

TEM images of the test structures showed that at the vulnerable interface of Cu, barrier metal and the dielectric, there is a noticeable change in the Cu profile and the density of the barrier metal. This can be assessed by the contrast in the image, which indicates that breakdown is likely to take place along this interface, as widely reported in the literature. Hence, the TF trend is caused by the change in linespace with linewidth, which in turn is caused by pattern dependency of etch.

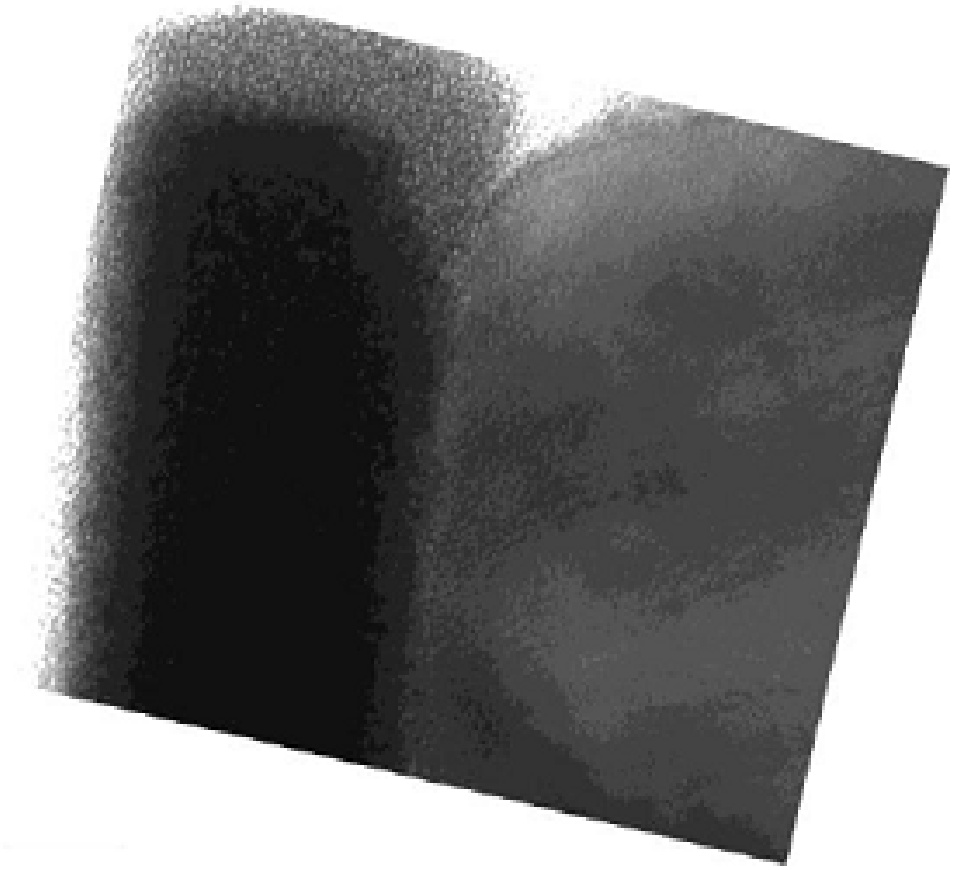


Figure 3.23: TEM image of Cu/dielectric/barrier metal interface. Darker colors indicate a denser material. Notice the change in color at the top of the barrier metal.

3.6 Modeling Characteristic Lifetime

Variation in linewidth, leading to variation in linespace, explains the observed difference in lifetime caused by ARDE.

3.6.1 Uniform Linewidth Test Structures

Characteristic lifetime is assumed to be a function of electric field in the dielectric. The electric field is a function of the distance between the lines. For a pitch, P, then the space, S_A, is

$$S_A = P - W_A. \tag{3.22}$$

Since,

$$P = W_D + S_D, \tag{3.23}$$

we have that

$$S_A = S_D - \Delta W. \tag{3.24}$$

The electric field is proportional to $1/s_A$, as noted in (3.1). Using (3.1) we have

$$\ln \eta = A + \frac{B}{S_D - \Delta W} \ , \tag{3.25}$$

for the E model, and

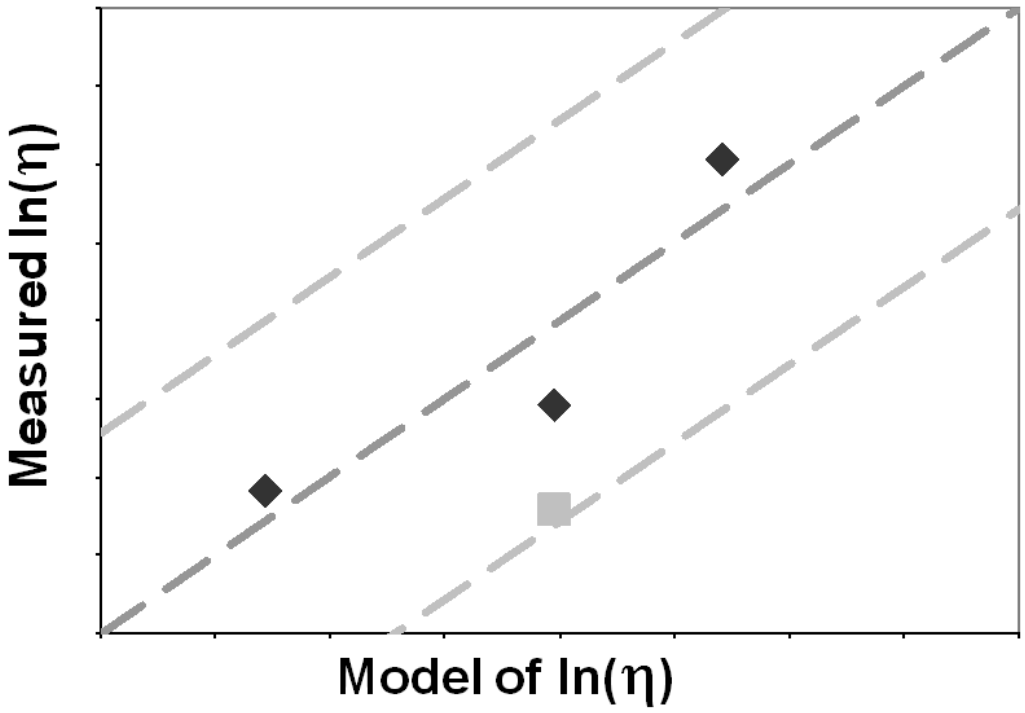
$$\ln \eta = A + \frac{B}{\sqrt{S_D - \Delta W}} \ , \tag{3.26}$$

for the $\sqrt{E}$ model.

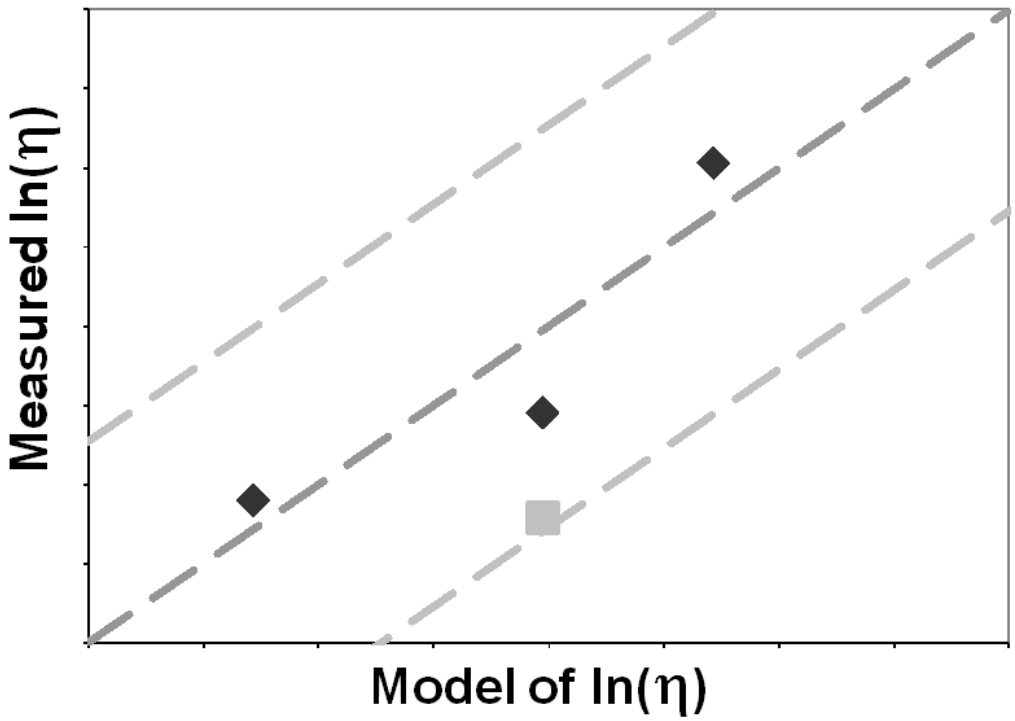
The characteristic lifetime data was used to find the best fit for A and B using only the data from the uniform structures. The results are shown in Figure 3.24.

3.6.2 Non-uniform Linewidth Test Structures

The models along with the computed constants, A and B, are used to predict lifetime for the non-uniform test structure, which has non-uniform actual linespace. This structure is equivalent to two structures in parallel, with the measured linespace, and



(a)



(b)

Figure 3.24: Model of the characteristic lifetime as a function of the measured value of characteristic lifetime using regression based on (a) $1/(S_D - \Delta W) = 1/S_A$ for the E model and (b) $1/\sqrt{S_D - \Delta W} = 1/\sqrt{S_A}$ for the $\sqrt{E}$ model. The 2σ confidence bounds are included. The black dots correspond to the uniform test structure, from which the model was computed, and the grey dots correspond to the non-uniform test structure.

half of the area. To do this it was noted that for the Weibull distribution, the cumulative probability density function is given by (3.9). From the Poisson distribution, the cumulative probability density function relates to defect generation as (3.10). Combining (3.9) and (3.10), a computed defect generation function is computed:

$$\lambda(TF) = \frac{1}{A}\left(\frac{TF}{\eta}\right)^{\beta}, \tag{3.27}$$

which produces $d(TF) = \lambda(TF)A$ as a function of time.

The $1X/5X$ test structure contains half of its vulnerable dielectric area with $1X$ linewidth on the left and $5X$ linewidth on the right, and it contains half of its vulnerable area with $5X$ linewidth on the left and $1X$ linewidth on the right. Therefore, its characteristic lifetime should be a combination of characteristic lifetimes of structures with these two vulnerable areas: $\eta_{1X,5X}$ and $\eta_{5X,1X}$. Let $\eta_{1X,5X}$ and $\eta_{5X,1X}$ correspond to defect densities $\lambda_{1X,5X}(TF)$ and $\lambda_{5X,1X}(TF)$, respectively. The combined defect generation function, for test structure $1X/5X$ is $d(TF) = \lambda_{1X,5X}(TF)A/2 + \lambda_{5X,1X}(TF)A/2$, i.e.,

$$d(TF) = \frac{1}{2}\left(\left(\frac{TF}{\eta_{1X,5X}}\right)^{\beta_{1X/5X}} + \left(\frac{TF}{\eta_{5X,1X}}\right)^{\beta_{1X/5X}}\right), \tag{3.28}$$

where A is the vulnerable area of the uniform test structures and $\beta_{1X/5X}$ is the shape parameter.

Substituting (3.28) into (3.9) gives the following joint cumulative probability density function:

$$P(TF) = 1 - \exp\left(-\frac{1}{2}\left(\left(\frac{TF}{\eta_{1X,5X}}\right)^{\beta_{1X/5X}} + \left(\frac{TF}{\eta_{5X,1X}}\right)^{\beta_{1X/5X}}\right)\right). \tag{3.29}$$

The characteristic lifetime of the joint test structure ($\eta_{1X,5X}$) is

$$P(TF) = 0.632. \tag{3.30}$$

Combining (3.29) and (3.30), we have

$$\left(\frac{\eta_{1X/5X}}{\eta_{1X,5X}}\right)^{\beta_{1X/5X}} + \left(\frac{\eta_{1X/5X}}{\eta_{5X,1X}}\right)^{\beta_{1X/5X}} = 2, \tag{3.31}$$

which results in

$$\eta_{1X/5X} = \left(\frac{1}{2} \left(\frac{1}{\eta_{1X,5X}^{\beta_{1X/5X}}} + \frac{1}{\eta_{5X,1X}^{\beta_{1X/5X}}} \right) \right)^{-1/\beta_{1X/5X}} . \tag{3.32}$$

There were no test structures corresponding to $\eta_{1X,5X}$ and $\eta_{5X,1X}$. Therefore, $\eta_{1X,5X}$ and $\eta_{5X,1X}$ were estimated with (3.25) and (3.26). The constants, A and B were determined with the uniform test structures.

As can be seen from (3.32), $\eta_{1X/5X}$ depends on $\beta_{1X/5X}$. It was found that β was a function of linewidth. It is best modeled as a function of the height of the lines, as shown in Figure 3.25.

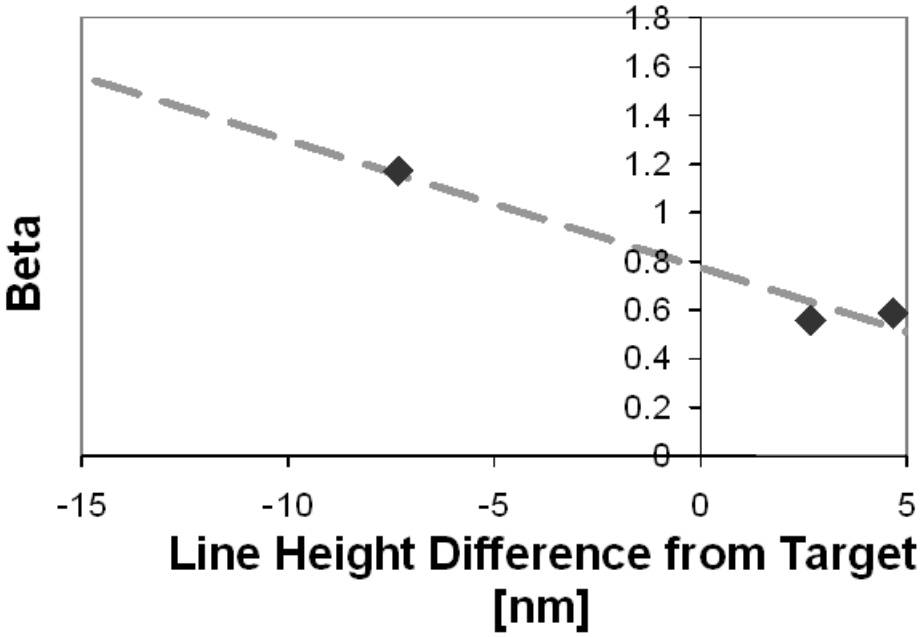


Figure 3.25: Variation in β as a function of line height. The model is fit with regression.

Therefore, $\beta_{1X/5X}$ is estimated by computing the normalized etch rate from Figure 3.14, as a function of the aspect ratios of the two lines. The normalized aspect ratio is converted to estimated line heights. The corresponding values of β were determined from Figure 3.25 and averaged. The results was $\beta_{1X/5X} = 0.81$. The value determined from the Weibull curve for the $1X/5X$ test structure was 0.83.

Figure 3.24 shows the predicted values of the characteristic lifetime based on (3.32) for both the E and $\sqrt{E}$ models. It can be seen that the model matches the data reasonably well.

3.6.3 Line Edge Roughness

Test structures showed line edge roughness (LER) as expected, with figures that are comparable to the ones reported in [139]. Line edge roughness changes the local linespace between interconnect lines and results in higher (and lower) local electric fields. Increased LER results in a reduction in the effective spacing, leading to a lower characteristic lifetime [66, 139]. Figure 3.26 shows a collection of SEM images of the top view of the test structures. In the absence of any LER the lines in the images would have been straight.

Figure 3.27 shows an example of the SEM image used to characterize LER. Sixteen different sites were characterized in the same way for the four linewidth test structures.

Equations (3.25) and (3.26) were used to assess the impact of LER on lifetime. LER is approximated as having a normal distribution with zero mean [66, 140]. Let S_i be the local linespace. Then S_i is distributed as $N(S_A, \sigma_{LER})$.

Let η be the characteristic lifetime associated with a test structure with vulnerable length, L, and linespace S_A. Let η_{LER} be the characteristic lifetime of a test structure with the same vulnerable length, L, impacted by LER. In this second test structure L is broken into segments such that $\sum_i L_i = L$, where each segment L_i is associated with line spacing $S_i \sim N(S_A, \sigma_{LER})$.

Let η_i be the lifetime of a test structure with vulnerable length L and linespace S_i. From (3.9) and (3.10), the total number of defects at breakdown for the test structure impacted by LER is

$$d(TF) = \sum_i d_i(TF) = \sum_i \frac{L_i}{L} \left(\frac{TF}{\eta_i(S_i)} \right)^{\beta}. \tag{3.33}$$

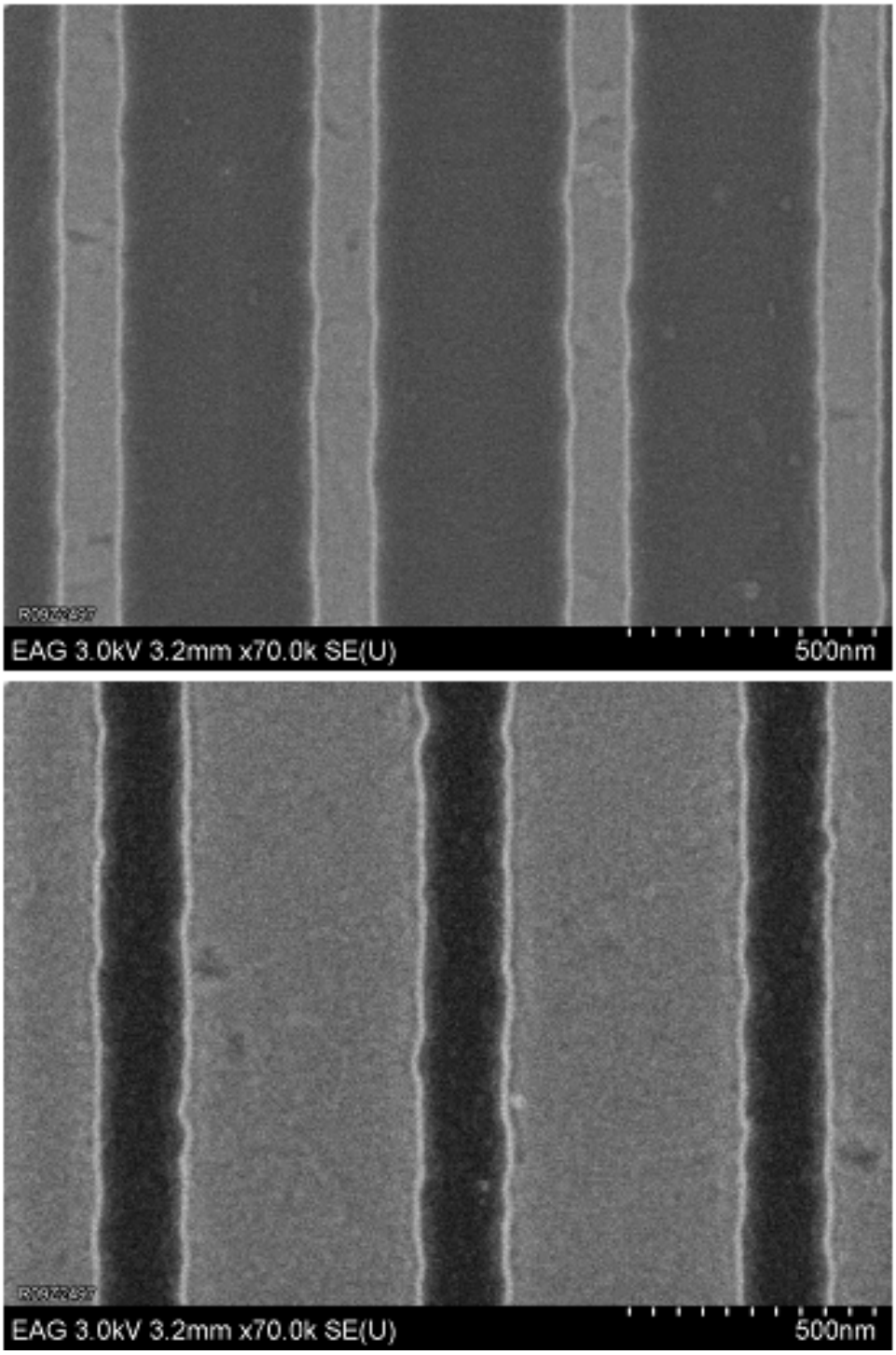


Figure 3.26: SEM images of the top view of the test structures showing LER. The images are from different sites on the uniform linewidth test structures, non-uniform linewidth test structure was also imaged but is not shown here.

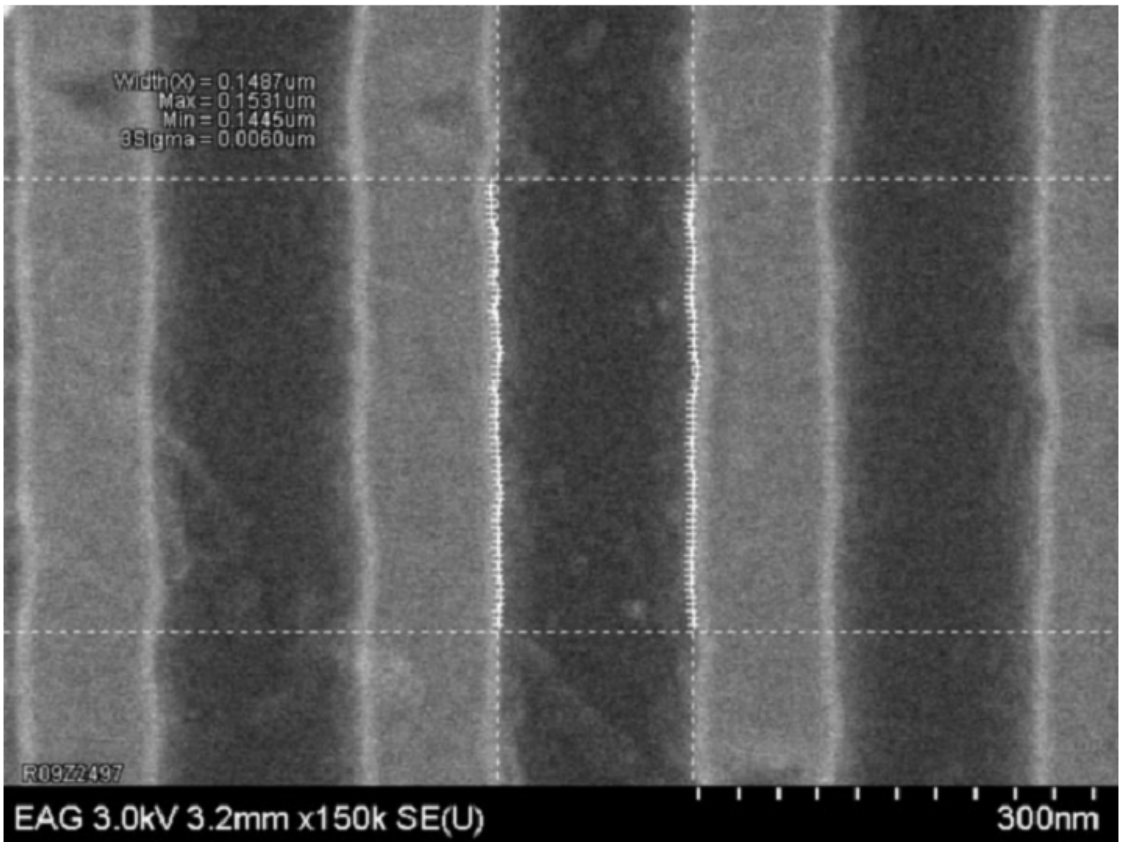


Figure 3.27: SEM images of the top view of the test structures used to characterize LER, four different sites on the four test structures were used for characterization. The upper left corner of the image shows the LER numbers for this site.

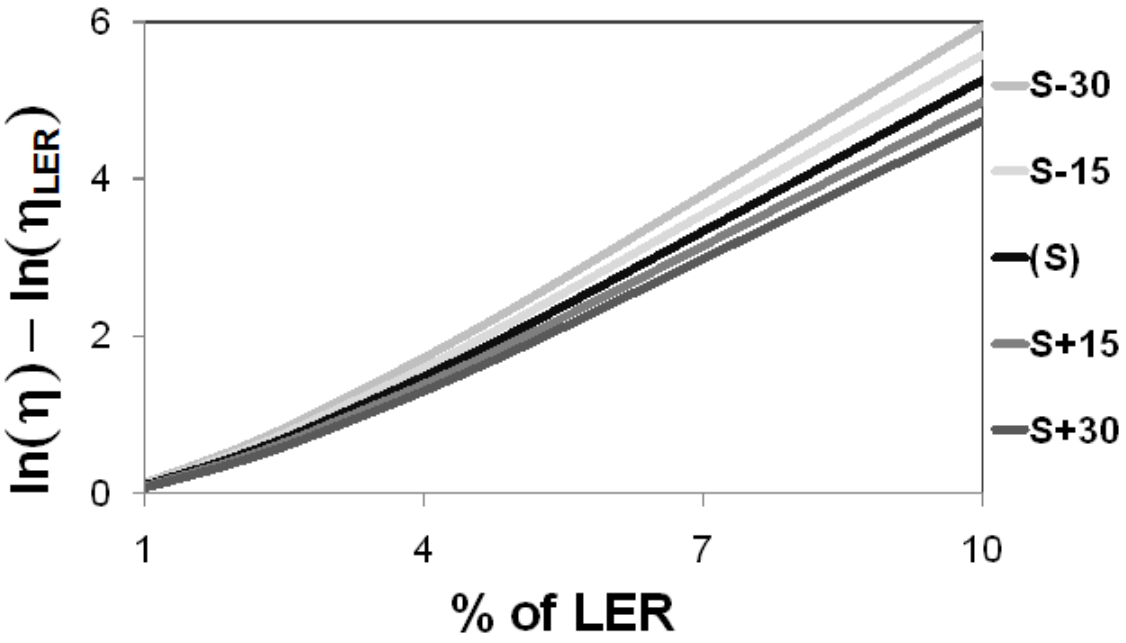
In the limit, for infinitesimally small segments

$$d(TF) = \int \left(\frac{TF}{\eta_i(S_i)} \right)^{\beta} f(S_i)\, dS_i \ , \tag{3.34}$$

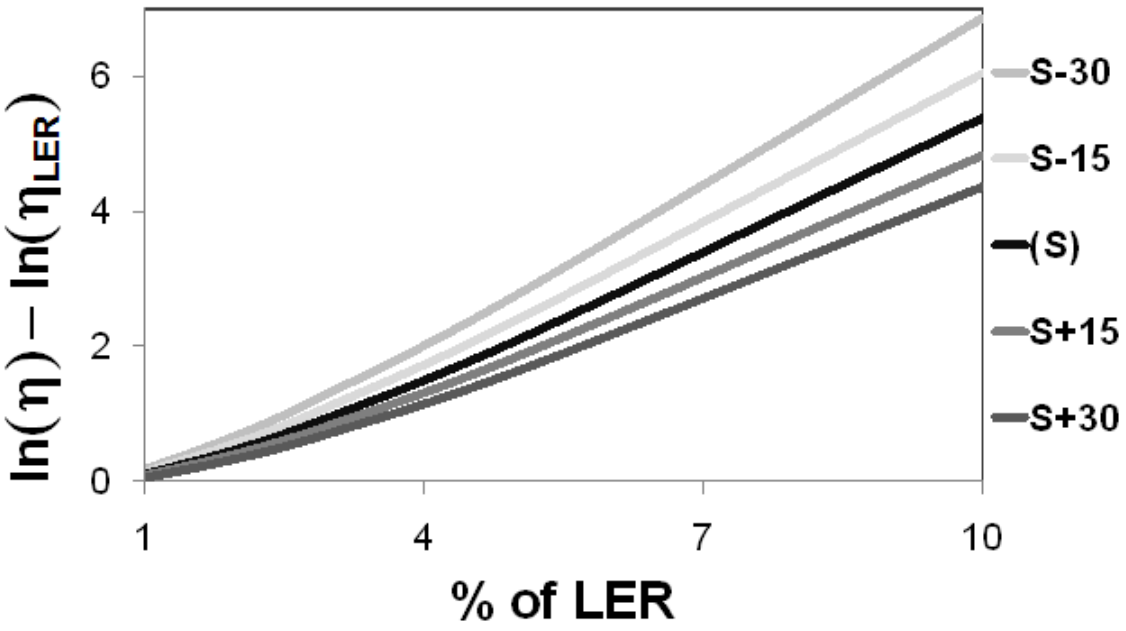
where $f(S_i)$ is a probability density function of a Normal distribution with mean S_A and variance σ^2_{LER}. A constant β is assumed on the basis that small changes in linespace cannot cause significant changes in β for a given applied voltage, although degradation in β has been reported in the literature at higher applied voltages [139].

The cumulative probability density function for the test structure with LER is $P(TF) = 1 - \exp\left(-d(TF)\right)$. Since the characteristic lifetime corresponds to $P(TF) = 0.63$, we have that

$$1 = \int \left(\frac{\eta_{LER}}{\eta_i(S_i)} \right)^{\beta} f(S_i)\, dS_i, \tag{3.35}$$



(a)



(b)

Figure 3.28: Shift in characteristic lifetime as a function of standard deviation of LER according to the (a) E model and (b) the $\sqrt{E}$ model.

or

$$\eta_{LER} = \left(\int \eta_i (S_i)^{-\beta} f(S_i)\ dS_i \right)^{-1/\beta}. \tag{3.36}$$

Equivalently

$$\eta_{LER} = \eta \left(\int \eta^{\beta} \eta_i (S_i)^{-\beta} f(S_i)\ dS_i \right)^{-1/\beta}. \tag{3.37}$$

Without loss of generality, let's consider the impact of LER on lifetime with the E model, i.e. Equation (3.25). In this vase, $\ln \eta = A + B/S_A$. Similarly, $\ln \eta_i = A + B/S_i$. As a result, for the E model

$$\eta_{LER} = \eta \left(\int \exp\left(\beta B \left(\frac{1}{S_A} - \frac{1}{S_i} \right) \right) f(S_i)\ dS_i \right)^{-1/\beta}. \tag{3.38}$$

Similarly, for the $\sqrt{E}$ model

$$\eta_{LER} = \eta \left(\int \exp\left(\beta B \left(\frac{1}{\sqrt{S_A}} - \frac{1}{\sqrt{S_i}} \right) \right) f(S_i)\ dS_i \right)^{-1/\beta}. \tag{3.39}$$

Figure 3.28 shows the results from Monte Carlo simulation. It shows the shift in characteristic lifetime as a function of the standard deviation of LER. As can be seen from the figure, the effect of LER is more pronounced at smaller linespaces, as noted in [66, 139]. As linespace increases, the shift in characteristic lifetime becomes progressively smaller.

For the used test structures, the linespace is sufficiently large that shift in $\ln \eta$ is approximately only 0.7 (in units of ln(seconds)). This is similar for all structures. The maximum difference in the impact of LER is 0.1 (in units of ln(seconds)). As a result, the prediction of the lifetime of the $1X/5X$ structure based on the $1X$, $3X$, and $5X$ structures (shown in Figure 3.24) is not affected.

3.7 Conclusion

Test structures were designed to model the impact of metal linewidth and pattern density on the behavior of backend low-k TDDB. The test structures varied both pattern density and linewidth independently to enable the separation of the impact

of these factors. TDDB was found to vary with linewidth, while pattern density had no effect on backend TDDB reliability. Different causes for the observed trends in lifetimes were considered. TDDB behavior was found to be dependent on aspect ratio dependent etching, which modulates the linewidths in printed structures and causes the distance between fingers of a comb structure to be non-uniform. Information on the relationship between linewidth and printed geometries on silicon was used to model the characteristic lifetime as a function of linespace. A good fit was obtained for the experimental data. Effect of LER on TDDB reliability was modeled and it was shown that LER impacts smaller linespaces more severely.

CHAPTER IV

BACKEND LOW-K TDDB CHIP RELIABILITY SIMULATOR

4.1 Motivation

Copper/low-k interconnects have proven to be the solution of choice to overcome issues raised by wire delays and dynamic power consumption in the deep submicron and nanometer regime. However, Cu/low-k interconnect systems are vulnerable to breakdown because of the nature of the materials and the complexity of the manufacturing process.

TDDB lifetime for different low-k dielectrics and conductor geometries is determined by using appropriate test structures. Lifetimes at use conditions are determined by scaling. To determine the lifetime of a chip, a correction is also needed to account for the difference between the vulnerable area of the product and the test structure. However, there is no literature on the method to find the vulnerable area for a chip for backend dielectric breakdown, except for the statement that the vulnerable area is "the total length of such [minimum spaced] lines within a product" [141].

The purpose of this framework is to clearly specify the vulnerable area of a chip for backend TDDB and to develop the link between data collected from test structures and chip lifetime due to backend dielectric breakdown. The proposed methodology accounts for more than minimum spaced lines and includes all layers of a chip, which can have different vulnerable areas.

This chapter begins with a brief background of electric field and temperature dependance of TDDB reliability. Section 4.3 highlights the need for a chip reliability simulator and gives a summary of different attempts of developing chip reliability

simulators. Section 4.4 formulates the problem and describes the approach to its solution. Section 4.5 summarizes test structure construction and test results. Section 4.6 outlines the methodology to determine backend low-k TDDB chip lifetime. Section 4.7 presents chip reliability results for the synthesized circuits. Section 4.8 details the impact of different layout optimizations on backend low-k TDDB chip reliability. Section 4.9 concludes the chapter.

4.2 Background

Several competing theories describe TDDB equally well [42, 53, 114]. Similarly, the effect of different conductor geometries on low-k TDDB has been the subject of various studies [57, 114]. Bulk of this work is built around the relationship between electric field (E) and TF [50, 142, 143].

The physics behind existing models cannot be easily extended to circuits. This is because lifetime is assumed to depend on only one circuit parameter. For instance, according to the E model,

$$TF = A_0 \exp\left(\frac{\Delta H_0 - \rho_{eff} E}{k_B T}\right), \tag{4.1}$$

where A_0 is a material constant, ΔH_0 is zero-field activation energy, k_B is the Boltzmann constant, and ρ_{eff} is the effective dipole moment and depends on the chemical composition of the dielectric. In (4.1), only temperature, T, and E impact lifetime. If it is assumed that electric field and temperature distributions are unrelated, then these models would require the computation of electric field intensity at the chip level. Such an analysis would require the finite element method to simulate the behavior of electric field in a chip, the computation of its average, together with a temperature profile simulator to compute the average temperature.

In general, models of backend TDDB are of the form

$$TF = k_1 \exp\left(-\gamma E^m\right), \tag{4.2}$$

where k_1 is a constant that depends on the material properties of the low-k dielectric, γ is a field acceleration factor that incorporates an inverse dependence on temperature, and $m = 1$ for the E model and $m = 1/2$ for the $\sqrt{E}$ model.

Reliability testing is accelerated by testing at higher electric fields and higher temperatures than use conditions. Time-to-failure is both a function of the electric field and the temperature. Equation (4.2) provides the correction that takes into account the difference between use condition and the electric field during test. The temperature dependence is modeled with an Arrhenius relationship [50]

$$\ln TF = B - \frac{C}{T}, \tag{4.3}$$

where B and C are a constant and a parameter respectively, for thermal processes. There is a concern that testing at high temperatures can activate failure modes that are not present during use conditions. Hence, testing at high electric fields is preferred in comparison with testing at high temperatures. Stress tests were conducted at 125°C. Equation (4.3) provides a correction between chip operating conditions and test conditions.

4.3 Full Chip Reliability Analysis

Chip reliability analysis requires techniques to extend the results gathered from small test structures and circuits to large complex chips. Such an endeavor must also be accompanied by solutions to manage and use the deluge of data that comes with analyzing large layouts. The physics describing IC failure mechanisms both in the front-end and in the backend has matured as a result of years of refinement to existing theories. However, the extension of these models to large and complex circuits has not proven to be straightforward and is complex.

The purpose of this chapter is to present a methodology to assess chip lifetimes based on low-k TDDB chip lifetimes, by developing the link between data collected from test structures and the chip. The feasibility of the proposed methodology is

demonstrated by presenting results from a simulator based on the proposed methodology. The methodology includes all layers of a chip, which can have different vulnerable areas.

The ultimate purpose of this work is to introduce backend dielectric reliability in design. The onus of meeting this end falls on the designers, if the reliability concerns and the accurate estimation of chip lifetimes can be conveyed to the designer in a designer-friendly manner. Designers ensure reliability, often inadvertently, by strictly adhering to design rules that assume worst case scenarios. Design rules often do not adapt themselves to the complexity of different circuits and operating conditions. Instead, design rules are kept general enough to encompass a large number of circuits. A design rule that may be too restrictive for one design can be completely unrealistic for another design.

As chip complexity increases, designers become less aware of the actual physical functioning of their chip. Similarly, extending the models of the physics of failure to a chip requires the consideration of a myriad of factors. The task can be simplified if test results are used incrementally, as proposed in this chapter.

4.3.1 Reliability Simulators

The most important reliability concerns for interconnect include electromigration, stress-induced voiding, and TDDB of the backend dielectric [141]. To date, reliability simulators for interconnect have only been developed for electromigration [144, 145]. There are both industrial and academic [145–147] simulators to assess reliability. However, they do not deal with the backend of the process. Moreover, those that deal with the backend of the process are limited to wearout caused by electromigration and do not take into account the effect of the dielectric on reliability [145]. The purpose of the proposed methodology is to consider a different wearout mechanism: backend TDDB, not currently included in reliability simulators. Hence, there are no

tools to predict the impact of low-k TDDB on lifetimes, even though low-k TDDB has emerged a major concern as Cu metallization has become mainstream.

It should be noted that circuits wearout for a variety of reasons, both related to devices and interconnect. All of these wearout mechanisms happen simultaneously. It is common to describe reliability mechanisms with a Weibull distribution with two parameters, the characteristic lifetime, η, and the shape parameter, β [117]. The characteristic lifetime is the time-to-failure at the 63% probability point, when 63% of the population has failed, and the shape parameter describes the dispersion of the failure rate population. Typically, the shape parameter is close to 1. If there is a collection of n wearout mechanisms, all of which are independent, modeled with Weibull distributions, with parameters, $\eta_i, i = 1, \ldots, n$, and $\beta_i, i = 1, \ldots, n$, then the characteristic lifetime of the system, η, i.e. the time when 63% of the population has failed from any mechanism is the solution of the following:

$$1 = \sum_{i=1}^{n} \left(\frac{\eta}{\eta_i} \right)^{\beta_i}. \tag{4.4}$$

If the characteristic lifetime of one mechanism is significantly smaller than others, this mechanism will dominate the failure rate. However, in general, it is prudent to consider *all* major sources of wearout. Hence, as the dimensions of the interconnect are lowered along with the dielectric constant, one should no longer neglect the potential reliability failure due to the backend dielectric.

4.4 Problem Formulation and Approach to the Solution

Given models for low-k TDDB, is there a way to predict chip lifetime because of low-k TDDB for different physical geometries and layout features, without directly using physics-based models?

4.4.1 Approach

To address the question above, first critical geometry features affecting backend TDDB reliability are identified. Test structure data is used to identify sensitivities of backend low-k TDDB to identifiable geometric features. The proposed framework focusses on backend TDDB. Although a number of factors can affect the TF of Cu/Low-k interconnects, in this work, as a starting point, focus is on the impact of area and metal linewidth on lifetime. The methodology is developed in a way that other reliability limiting mechanisms can be integrated in the methodology. Models are built to extrapolate failure times from one feature size to another [122]. These geometric features are then used to extract the vulnerable area of chips and to characterize failure rates [64, 122]. And, models are used to combine failure rates from different features to extract a full chip failure rate.

This work forms an interface among reliability physicists, semiconductor foundry engineers, and designers by suitably partitioning their combined effort, thereby keeping each unburdened with the details of the other's efforts. Test structures have identified the well known sensitivities of backend dielectric breakdown to area and distance between the lines (electric field in the dielectric) and the less well known sensitivity to metal linewidth. Through the methodology, test structure data is linked to the entire product die (chips), by extracting the corresponding "vulnerable areas" for a chip, and the failure rate for the chip is characterized by combining the failure rates due to all "vulnerable areas".

The simulator combines the test results from test structures that stress different vulnerable areas to determine the failure rate for a chip. In other words, the methodology links chip layout geometries to those on the test structures. It is assumed that these results can be scaled to use conditions with whatever version of Equations (4.1) and (4.2) that is deemed to be appropriate.

The focal point of this work is the mapping between the test structures and the

chip. It is assumed that the conductors in the chip are very similar to the conductors in the test structures and that they have similar geometries. Conductor geometry does impact the data that is collected from the test structures and the models developed from these data. The methods that account for conductor geometry will be noted in the following sections.

4.5 Test Structure Design and Test Results

4.5.1 Test Structures

Test structures were designed to assess the impact of area and linewidth on Cu/low-k TDDB. The details of the test structures, their design and results, are given in Chapters 2 and 3 [64, 122]. The test structure set includes comb structures that vary area, number of tips, linespace, and linewidth. Test structure design and test results are summarized below.

The test structures were manufactured with an industrial $45nm$ dual-damascene process and subjected to accelerated stress tests at 125°C with electric fields ranging from $0.25MV/cm$ to $1.5MV/cm$ Breakdown was considered as the point of onset of leakage current greater than $100\mu A$. The sample size was 30. Weibull failure rate distribution was used to model the failure population.

4.5.2 Test Results

Test results indicated a strong impact of area. Die-to-die linewidth variation creates curvature in failure rate distributions. This curvature does not impact η. Hence η is extracted and β is determined by area scaling. Once these parameters have been determined for the unit area, the relationship between characteristic lifetimes for different areas in known.

Note that LER can also be taken into account when extrapolating failure rates. The impact of LER was considered in Chapter 3 and [64].

Lifetime is also impacted by the linewidth on each side of the dielectric segment.

Linewidth variation was considered when determining η and β for the test structures, in Chapter 2 [122]. Specifically, linewidth can be taken into account by extracting η as a function of linewidth since β, which is extracted from the area test structures, is assumed to be constant.

Test results showed a strong impact of linewidth on low-k TDDB, even when the vulnerable area and linespace of the dielectric under stress remained constant. If W_A is the actual linewidth and W_D is the drawn linewidth, then the difference between them is given by $\Delta W = W_A - W_D$. The shift in linespace is $S_A = S_D - \Delta W$, where S_A is the actual linespace and S_D is the drawn linespace. This shift arises because of aspect-ratio-dependent-etching (ARDE) [129]. During etch a protective compound builds up on the sidewalls of wide trenches, preventing lateral etch. This compound does not build up as much in narrow trenches. Hence, narrow trenches suffer from greater lateral etch near the critical CMP interface. Linespaces with larger positive values of ΔW breakdown faster, since $E = V/S_A$. Data can be used directly to determine the relationship between the drawn linespace and η through regression.

Data indicated a difference in linespace as a function of the widths of lines on the right and left as shown in Figure 4.1.

If SEM data is available, then Equation (4.2) along with $E = V/S_A$ can be used to calculate k_1 and γ. The ΔW can also be found for any block of dielectric in the layout with any linewidth on the right and the left. The model in Figure 4.1 gives an estimate of ΔW and S_A. The constants from Equation (4.2), in turn, provide an estimate of the corresponding characteristic lifetime.

Figure 4.2 shows a plot of characteristic lifetime varying with area and linespace, extracted from the test structures. This plot is obtained by determining the characteristic lifetime for different area ratios, in comparison with the $1X$ test structure, for different linespaces, i.e.,

$$\ln \eta \propto f(S, A),$$

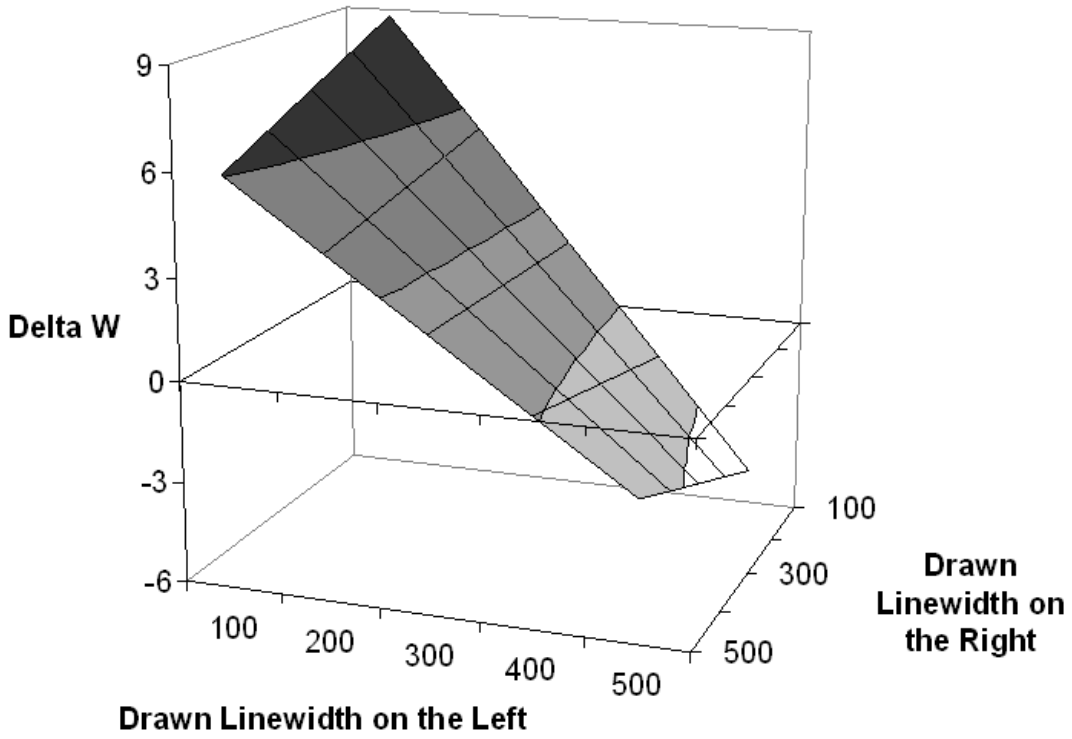


Figure 4.1: Variation in linespace as a function of the widths of the lines in either side of the dielectric. The data was collected using scanning electron microscopy.

where S is the linespace and A is the area.

4.5.3 Scaling Test Results to Use Conditions

Electric field and temperature can affect the relationship between test conditions and use conditions. The relationship between test conditions and use conditions is given in Equation (4.2). However, the test structure is stressed with constant voltage stress (CVS), DC stress, while the chip dielectrics undergo AC stress. Nonetheless, it should be noted that the backend dielectric TDDB under AC stress does not show any recovery [68], as observed in bias temperature instability degradation, and lifetime relaxation or healing, as observed in degradation due to electromigration.

A signal activity factor of 0.5 was assumed. Figure 4.3 shows stress test results scaled to use conditions for $45nm$ technology, with a supply voltage of $0.8V$ under

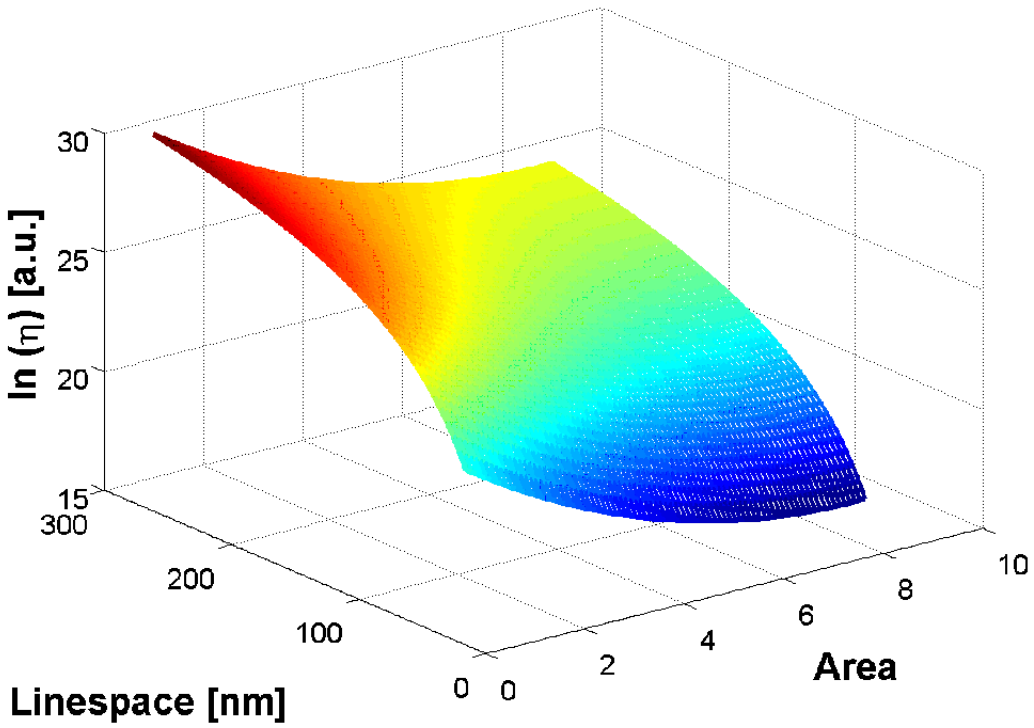


Figure 4.2: The effect of dielectric area and linespace on Weibull characteristic lifetime (η) using the $\sqrt{E}$ model. The model is based on data taken from test structures, where linespace is determined by linewidth, since the pitch is constant. Area is the ratio of the area of extrapolation to the unit area test structure.

alternating pulsed stress.

It should be noted that segments of the circuit may undergo different signal activity factors. In that case, these sectors should be analyzed separately, before combining them with estimates of lifetime of other sectors. Similarly, segments of the circuit may experience different values of average temperature. These sectors should also be analyzed separately, before combination with estimates of lifetime from other sectors.

4.6 Lifetime from Chip Layout

4.6.1 Vulnerable Sites

Based on the information that can be extracted from a layout in GDSII format, sites that are vulnerable to low-k TDDB, thus the vulnerable area, are determined. A

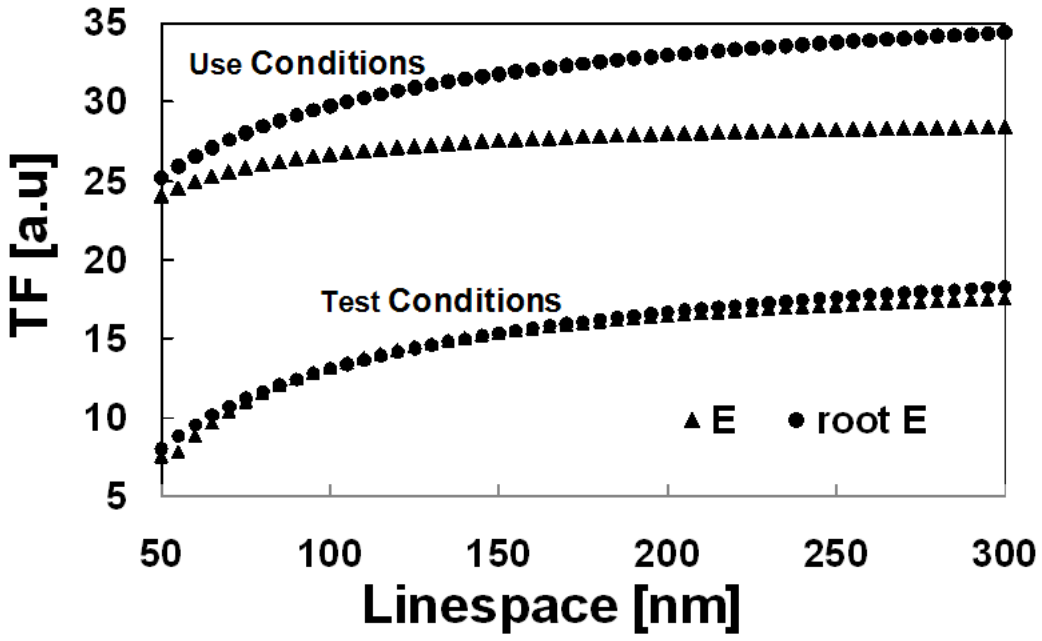


Figure 4.3: Test results scaled to use conditions.

vulnerable area is a two dimensional block of dielectric, with a horizontal (linespace) and a vertical dimension (length). The feature that is extracted from layout is the vulnerable length (L) between two lines associated with a linespace (S), which is a function of the two adjacent linewidths (W_L, W_R) [64, 148–150]. The vulnerable length L_i surrounded by lines separated by linespace S_i is illustrated in Figure 4.4. The layout is analyzed by determining the pairs $(S_i(W_L, W_R), L_i)$ for each layer, for all linespaces $S_i(W_L, W_R)$. If the different areas in the layout have different average on-chip temperatures, then the layout can be partitioned into sectors according to the temperature.

4.6.2 Characteristic Lifetime for a Layer

A Poisson model is assumed for the defect distribution over an area and the Weibull failure distribution

$$P = 1 - \exp\left(-\lambda(t)A_f\right) = 1 - \exp\left(-\left(\frac{TF}{\eta}\right)^{\beta}\right), \tag{4.5}$$

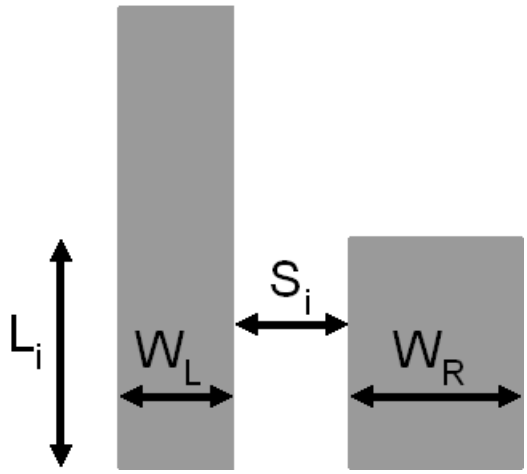


Figure 4.4: The vulnerable length, L_i, is the length for which two lines run side by side separated by linespace S_i.

where $d_f = \lambda(t)A_f$ is the number of defects at failure for the f^{th} feature, $\lambda(t)$ is the defect density at failure, and A_f is the vulnerable area. Let $A_{test} = S_{test}L_{test}$ be the vulnerable area of the test structure, whose actual line spacing is S_{test}, which corresponds a vulnerable length, L_{test}. Then, the test data is used to find $\lambda(t)$ corresponding to S_{test}.

Suppose that the vulnerable length corresponding to actual spacing S_f is L_f. Then the corresponding number of defects for the spacing S_f is $d_f = \lambda(t)S_fL_f$. Since the characteristic lifetime corresponds to $P = 0.63$, the corresponding characteristic lifetime of the chip is related to the lifetime of the test structure as

$$\eta_f = \eta_{test}\left(\frac{L_{test}}{L_f}\right)^{1/\beta}. \tag{4.6}$$

Note that the layout contains more than one spacing between the lines. Therefore, how to determine a lifetime when the line spacing is different than the test structure must be considered. To do this, (4.2) is used and rewritten in terms of [148–150]

$$\eta_{test} = k_1 \exp\left(\frac{-\gamma V^m}{S_{test}^m}\right). \tag{4.7}$$

Now, let us suppose that the chip contains line spacing S_f. Then, the corresponding characteristic lifetime of a chip with line spacing S_f and vulnerable length L_f

is

$$\eta = \eta_{test} \exp\left(-\gamma V^m \left(S_f^{-m} - S_{test}^{-m}\right)\right) \left(\frac{L_{test}}{L_f}\right)^{1/\beta}. \tag{4.8}$$

Layouts contain many different line spacings, S_f. Hence, it is necessary to combine the failures for all lines spacings. This is done by computing an overall defect count at failure. For a single spacing, the defect count, d_f, as a function of time is

$$d_f = \frac{L_f}{L_{test}} \left(\frac{TF}{\eta_{test}}\right)^{\beta} \exp\left(\gamma V^m \beta \left(S_f^{-m} - S_{test}^{-m}\right)\right). \tag{4.9}$$

The total defect count at failure is the sum $d = \sum_f d_f$, which is the sum of all defect counts of all linespaces present in a layer. This cumulative defect count is used to find the characteristic lifetime of the layer (η_{layer}), at probability point $P_i = 0.63$, implicitly defined as the solution to

$$1 = \sum_f \left(\frac{\eta}{\eta_f^{\beta}}\right)^{\beta_f}.$$

In the limit when β_f is constant [148, 150],

$$\eta_{layer} = \left(\sum_f \frac{1}{\eta_f^{\beta}}\right)^{-1/\beta}, \tag{4.10}$$

which is equivalent to [148–150]

$$\eta_{layer} = \eta_{test} \left(\frac{L_{test}}{\sum_f L_f \exp\left(\gamma V^m \beta \left(S_f^{-m} - S_{test}^{-m}\right)\right)}\right)^{1/\beta}. \tag{4.11}$$

Figure 4.5 shows the characteristic lifetime for each layer for an example JPEG encoder/decoder chip. The lifetime of layer 3 is shortest because it is the densest.

4.6.3 Characteristic Lifetime for Chip

Let d_l be the defect density for each layer, where

$$d_l = \sum_f d_f, \tag{4.12}$$

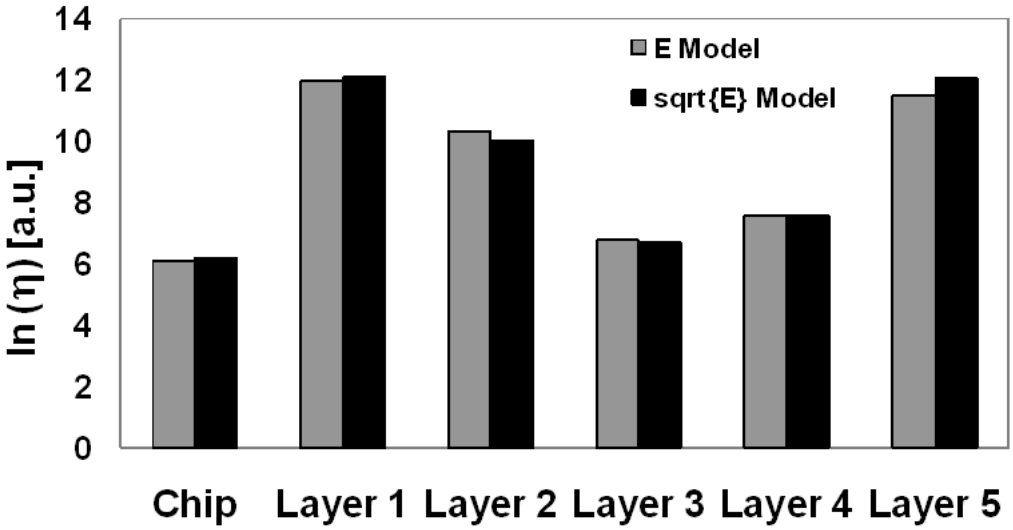


Figure 4.5: Characteristic lifetime for individual layers and the complete chip for a JPEG decoder/encoder. The figure also shows the most frequent line spacing for each layer.

and is computed with Equation (4.9). Overall, for the chip,

$$d_{Chip} = \sum_{l} d_l. \tag{4.13}$$

Unlike for a single layer, multiple layers of a chip may have different process details. Therefore, data would need to be collected from test structures for each layer separately, i.e., L_{test}, η_{test}, and $\beta(l)$ are unique to each layer.

If β were common to all layers, then it is possible to solve for the characteristic lifetime of the chip, η_{Chip}, as

$$\eta_{Chip} = \left(\sum_{l} \sum_{f(l)} \eta_{f(l)}^{-\beta} \right)^{-1/\beta}, \tag{4.14}$$

which is equivalent to

$$\eta_{Chip} = \left(\sum_{l} \frac{\sum\limits_{f(l)} L_{f(l)} \exp\left(\gamma V^m \beta \left(S_{f(l)}^{-m} - S_{test(l)}^{-m}\right)\right)}{L_{test} \left(\eta_{test(l)}\right)^{\beta}} \right)^{-1/\beta}. \tag{4.15}$$

Otherwise, η_{Chip} is implicitly defined as

$$1 = \sum_{l} \sum_{f(l)} \left(\frac{\eta_{Chip}}{\eta_{f(l)}} \right)^{\beta_{f(l)}}, \tag{4.16}$$

i.e.,

$$\sum_{l} d_l = 1, \tag{4.17}$$

where [64, 148–150]

$$d_l = \sum_{f(l)} \frac{L_{f(l)}}{L_{test(l)}} \left(\frac{\eta_{Chip}}{\eta_{test(l)}} \right)^{\beta(l)} \exp\left(\gamma V^m \beta(l) \left(S_{f(l)}^{-m} - S_{test(l)}^{-m} \right) \right). \tag{4.18}$$

Projection of time-to-failure at small probabilities, P, requires not just η_{Chip}, but also β_{Chip}, where β_{Chip} is

$$\beta_{Chip} = \left. \frac{\partial(\ln(-\ln(1-P)))}{\partial(\ln(t))} \right|_{t=\eta_{Chip}}. \tag{4.19}$$

P is the cumulative probability density function of failure for the chip. Solving [148],

$$\beta_{Chip} = \sum_{l} \sum_{f(l)} \beta_{f(l)} \left(\frac{\eta_{Chip}}{\eta_{f(l)}} \right)^{\beta_{f(l)}}. \tag{4.20}$$

To compute the lifetime at probability point, P, say, $P = 0.0001$, the lifetime, t, is

$$t = \eta_{Chip} \left(-\ln(1-P) \right)^{1/\beta_{Chip}}. \tag{4.21}$$

Since the test structures only give data from one layer, it was assumed that CMP, etching and photolithography impact all the layers in the same way. This assumption is simplistic, and if data from different layers is available, it can be easily incorporated into Equations (4.10)–(4.16).

Reliability is adversely affected by linewidth variation and LER. It has been shown that large scale linewidth variation impacts β [122], while LER impacts η with little or no impact on β [65]. Linewidth variation is considered when determining β for the test structures [64]. Any effect of LER is reflected in the characteristic lifetime through η_{test} in Equations (4.10) and (4.16) [65], and thus is included in the results.

4.6.4 Temperature Profile

Including the temperature map in the layout statistics adds another dimension to the problem because different characteristic lifetimes at different temperatures for every linespace must be considered now. For m different temperature regions for a linespace S_1, then the corresponding characteristic lifetime for the linespace is

$$\eta_{S_1} = \sum_m \left(\frac{1}{\eta_m^\beta} \right)^{-1/\beta}. \tag{4.22}$$

Characteristic lifetimes for the layer and the chip can be calculated using Equations (4.10) and (4.16).

4.7 Low-k TDDB Chip Reliability

4.7.1 Experimental Setting

4.7.1.1 Process

NCSU 45nm technology library was used for synthesizing circuits used in simulations [151]. This process has ten metal layers and the details of relevant features are given in Table 1.

Table 1: Metal layers in NCSU 45nm technology

Metal Layer	Minimum Linewidth $[nm]$	Minimum Linespace $[nm]$
1	65	65
2, 3	70	70
4, 5, 6	140	140
7, 8	400	400
9, 10	800	800

4.7.1.2 Circuits

Radix-2 pipelined, 256 points and 512 points, fast fourier transform (FFT) HDL source code, downloadable from [152], was synthesized. The circuit *cf_fft_256_8* has

$324k$ gates and $329k$ nets. The circuit *cf_fft_512_8* has $708k$ gates and $712k$ nets. Both circuits have precision 8. Names of different instantiations of *cf_fft_256_8* and *cf_fft_512_8* start with $f1$ and $f2$, respectively. The block diagram of the circuit is shown in Figure 4.6.

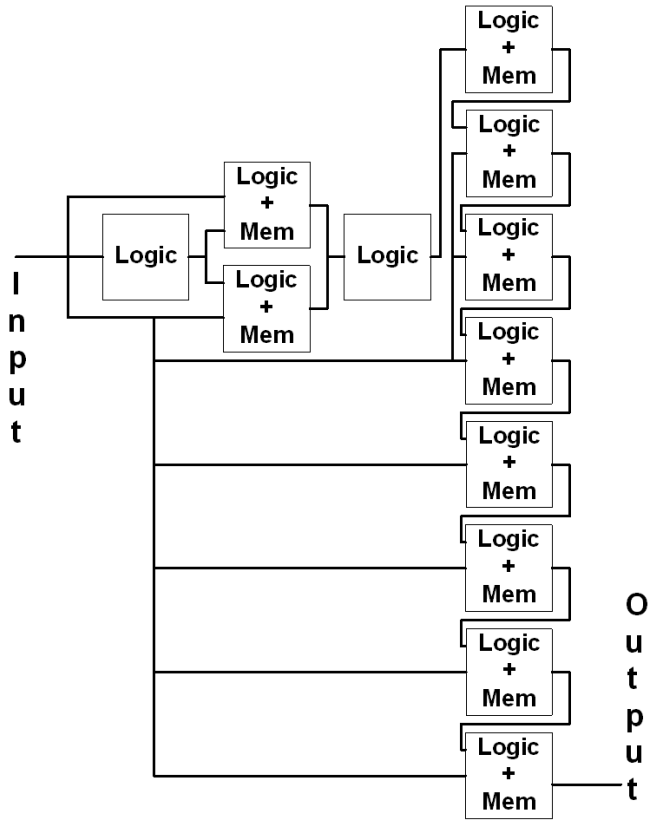


Figure 4.6: Block diagram of FFT circuit used for reliability evaluation.

Synopsys Design Compiler was used for synthesis [153]. Cadence SoC Encounter was used for placement, clock-tree synthesis, routing, optimization, and RC extraction [154]. Synospys PrimeTime was used for timing analysis [155]. Seven different instantiations of *cf_fft_256_8* and four different instantiations of *cf_fft_512_8* were used.

The metrics of performance comparison were the number of layers in a circuit and its timing performance. The details of circuits are given in Table 2. Table 2 shows the timing performance of each circuit, the total wirelength of each circuit, and the percentage of total wirelength in each layer, referred to as wire density.

Table 2: Wirelength of individual metal layers, as well as timing performance and reliability, of the designs used. Table shows the percentage of total wirelengths of chip present in each layer. Wirelength (WL) and timing performance, critical path delay (CPD), are also given.

Layer	Design									
	f1_M5	f1_M6	f1_M7	f1_M8	f1_RT1	f1_RT2	f1_RT3	f2_RT1	f2_RT2	f2_RT3
Metal1	1.4	1.4	1.4	1.4	2.6	0.8	2.2	2.2	2.2	1.2
Metal2	18.8	18	18.1	18.1	22.6	14.1	19.5	20.8	23.3	15.8
Metal3	33.9	33.1	33.1	33.1	38.4	25.8	32	35.4	37.9	27.9
Metal4	29.5	25.7	25.7	25.6	24.9	18.9	23.1	24.9	25.3	22.7
Metal5	16.2	16.1	15.4	15.3	9.52	19.9	15.5	13.4	9.04	19.8
Metal6		5.4	5.07	5.03	1.86	15.9	7	2.77	2.07	12.53
Metal7			0.9	1	0.05	4.6	0.7	0.38	0.04	
Metal8				0.1		0.01			0.004	
Total WL [m]	8.06	8.05	7.99	7.99	6.28	13.56	7.52	15.06	13.91	21.09
CPD [ns]	3.51	3.51	3.33	3.29	3.16	2.9	2.86	2.91	2.96	2.98

Circuits $f1\_M5$, $f1\_M6$, $f1\_M7$, and $f1\_M8$ were used to isolate the impact of number of layers on reliability. Circuits labeled 'MX' use Metal1 to Metal'X' during routing. Using more routing layers tends to have shorter wirelength and better timing performance, as shown in Table 2.

Circuits $f1\_RT1$, $f1\_RT2$, $f1\_RT3$, $f2\_RT1$, $f2\_RT2$, and $f2\_RT3$ were used to analyze the impact of timing performance on reliability. In RT'Y', timing optimizations using buffer insertion and gate sizing was used, M'X' does not do this. A higher value of 'Y' means more aggressive timing optimization with higher clock frequency.

All the circuits were synthesized using the same technology library, thus the values in Table 1 were consistent across all the instantiated circuits. The details of vulnerable area extraction can be found in [156].

4.7.2 Runtime

The runtime for the simulator is the sum of the time taken to extract features from layout and a constant time to evaluate Equations (4.9)–(4.16). Complexity of feature extraction and database extraction is $O(n)$, where n is the number of features, since bucket-sort is used. Complexity of extracting statistics from features is also $O(n)$ because the bucket is scanned from the bottom most element, and the maximum number of features within a fixed distance from element is constant. Lifetime is estimated in constant time.

4.7.3 Results

Figure 4.7 shows η for Metal1–Metal6 for the circuits used in the study and the chip according to the $\sqrt{E}$ Model. η for chips are more pessimistic than η for individual layers because in calculating characteristic lifetime for the chip vulnerable areas across all the layers are combined. Figure 4.7 does not indicate a trend for reliability against timing performance, however it shows that increasing the number of layers affects reliability marginally while decreasing wirelengths increases reliability.

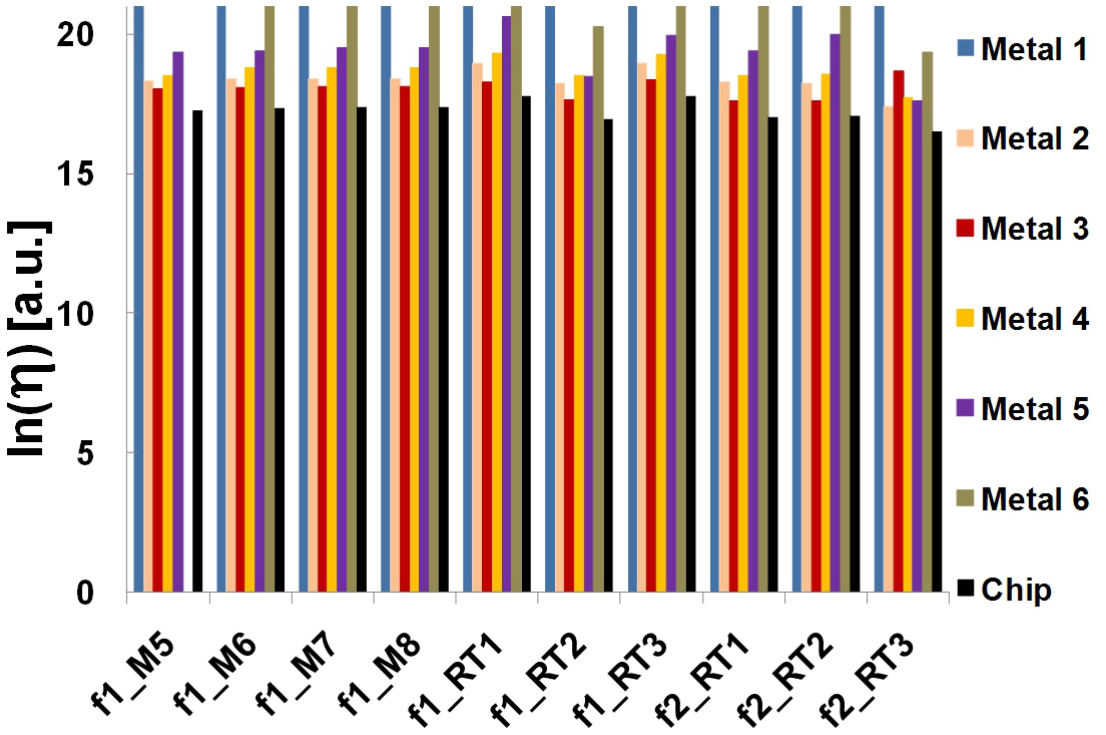


Figure 4.7: Characteristic lifetimes for each layer and the chip using the $\sqrt{E}$ model.

Figure 4.8 shows characteristic lifetimes for the circuits simulated according to the E model and the $\sqrt{E}$ Model. As noted previously, the E model gives a more pessimistic lifetime estimate of the two models [48].

4.8 Impact of Layout on TDDB Reliability

4.8.1 Observations

Characteristic lifetimes for chips are more pessimistic than η for individual layers, because in calculating characteristic lifetime for the chip vulnerable areas for all the layers are combined. Figure 4.7 does not indicate a trend for reliability with respect to timing performance. Figure 4.9 shows the lack of correlation between timing performance and reliability. Results showed that increasing the number of layers affects reliability marginally, while decreasing wirelengths increases reliability, as shown in Figure 4.9.

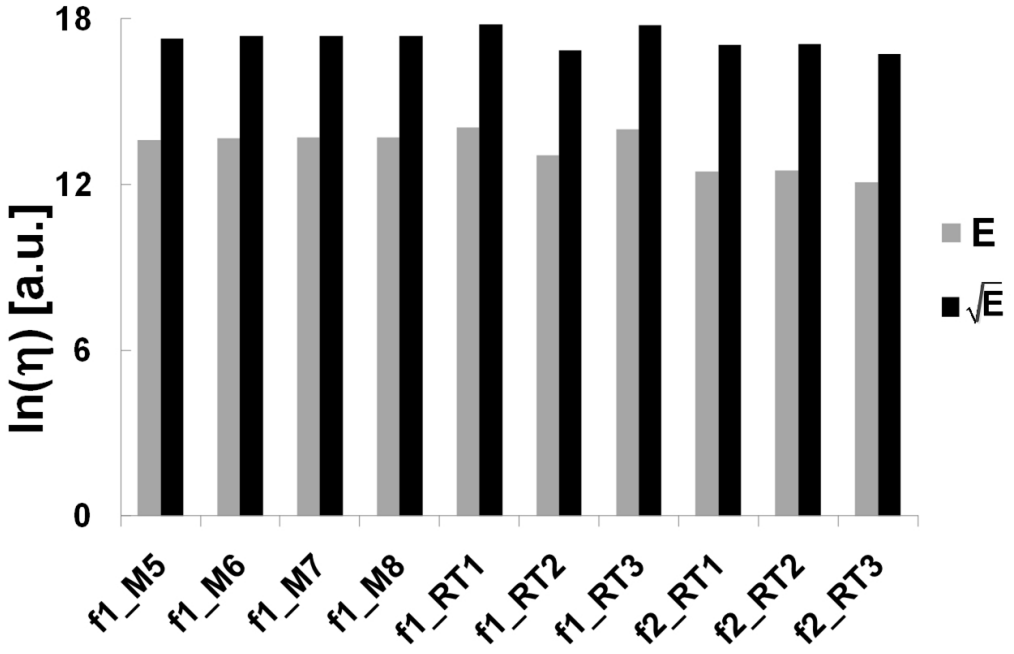


Figure 4.8: Characteristic lifetime for circuits according to the E Model and the $\sqrt{E}$ Model.

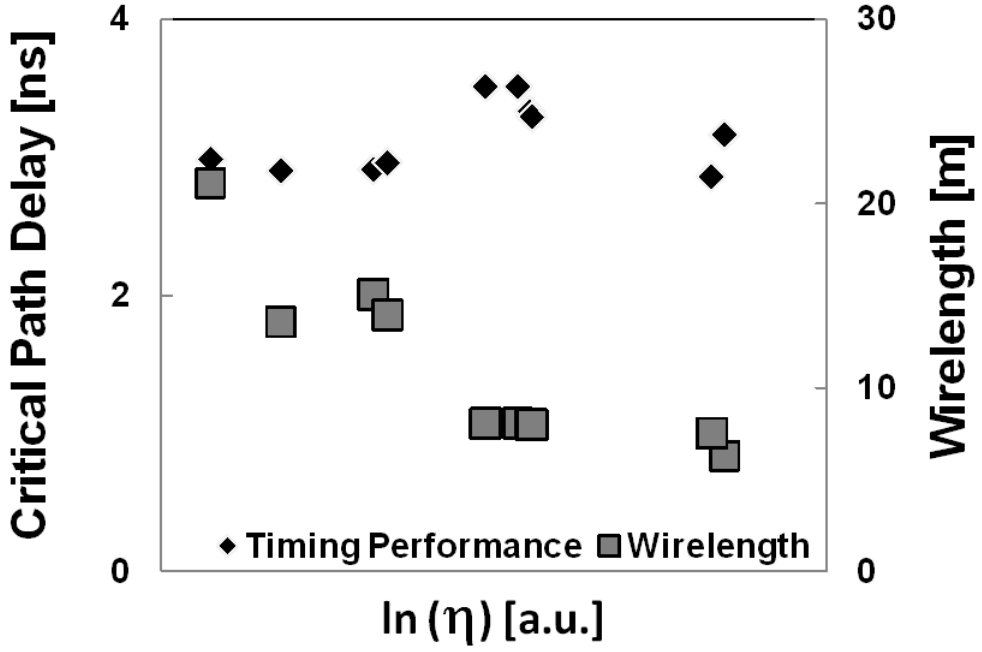


Figure 4.9: A comparison of reliability, timing performance and wirelength for the circuits under study.

4.8.1.1 Number of Layers and Reliability

Using more metal layers generally results in a decrease in routing congestion and obviates the need of long detours for avoiding routing congestions. This leads to less coupling capacitances between wires resulting in a decrease in the critical path delay. Since a router can spread out wires in several metal layers, this is expected to improve reliability.

As expected, the critical path delay goes down as the number of metal layers is increased. However, as shown in Figure 4.7, reliability increases only marginally as the number of layers increase, although the layer critical to lifetime remains the one with highest wire density. This change, or lack thereof, can be expected because even though the number of layers increase from 5 to 8, the percentage of total wirelength in additional layers is less than 6%. Not only that, but also large percentage of the wirelengths still remain in a single layer, Metal3. Metal3 has mid-distance interconnects performing vital operations and it is highly unlikely that any particular optimization would cause major changes in Metal3. An even distribution of wirelengths can lead to an increase in lifetime.

4.8.1.2 Critical Physical Features for TDDB

All dielectric area in a layer and in the chip falls in some linespace group, determined by its immediate neighbors. According to Equation (4.10), η for a layer is dominated by the η of the critical linespace group, i.e., the most frequent linespace group. Figure 4.10 shows the percentage of dielectric area formed by the critical linespace group. A way to increase η for layers with a critical linespace greater than the smallest linespace is to redistribute linewidths. This distribution can be optimized for reliability.

If the characteristic lifetime was to be estimated based on the most frequent (critical) linespacing alone, only the area for this single linespace in each layer needs

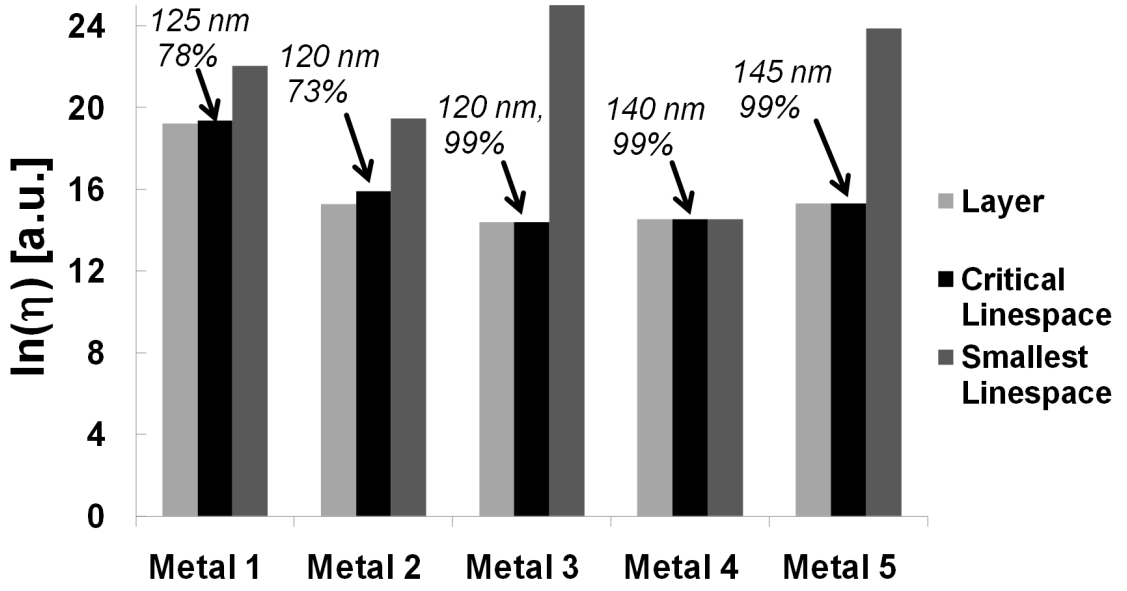


Figure 4.10: Characteristic lifetime for individual layers for, a radix-2 pipelined FFT 8 chip. The figure also shows the most frequent line spacing for each layer and the percentage of vulnerable area covered by this most frequent line spacing. It compares the estimated lifetime including all vulnerable areas, the vulnerable area associated with only the most frequent vulnerable area (critical line space), and the vulnerable area associated with only the smallest line space.

to be determined. Such an approach is simplistic. However, Figure 4.10 shows that lifetime estimates based on the critical linespace group are reasonably accurate.

Figure 4.10 also shows the characteristic lifetime for the minimum linespace group for each layer, where the minimum linespace for each layer is given in Table 1. Consider Metal2 in $f1\_M5$, the smallest linespace in Metal2 is $70nm$, but the linespace dominating the characteristic lifetime for this particular layer is $120nm$ for both the E Model and the $\sqrt{E}$ Model. Forty different linespaces were present throughout the layer, with the minimum being $70nm$ and the maximum being $252.5nm$. However, 73% of the dielectric segments in this layer had a linespace of $120nm$, with only 0.11% of dielectric segments having a linespace of $70nm$ between them. Therefore, the smallest linespace group just cannot be considered, as suggested in [141], when the layout is dominated by a linespace group other than the minimum linespace. Such

an approach leads to lifetimes that are optimistic by orders of magnitude.

4.8.1.3 Process Variation and Reliability

Reliability is adversely affected by linewidth variation and LER. It has been shown that large scale linewidth variation impacts the slope of the Weibull Curve, β, [123], while LER impacts characteristic lifetime, η, with little or no impact on β [65]. Linewidth variation was considered while determining η and β for the test structures [123]. Any effect of LER is reflected in η_{test} in equations (4.10) and (4.16) [65].

4.8.1.4 Wire Density and Reliability

Figure 4.11 shows that there is a strong correlation between wire density and backend TDDB reliability, with layers having the highest wire density dominating η in (4.16). This is an expected result because of the nature of the breakdown mechanism. Higher wire coverage is achieved by closely packing the metal lines together, resulting in an increase in E and consequently degrading reliability. Also, as expected, reliability increased with a decrease in wirelength.

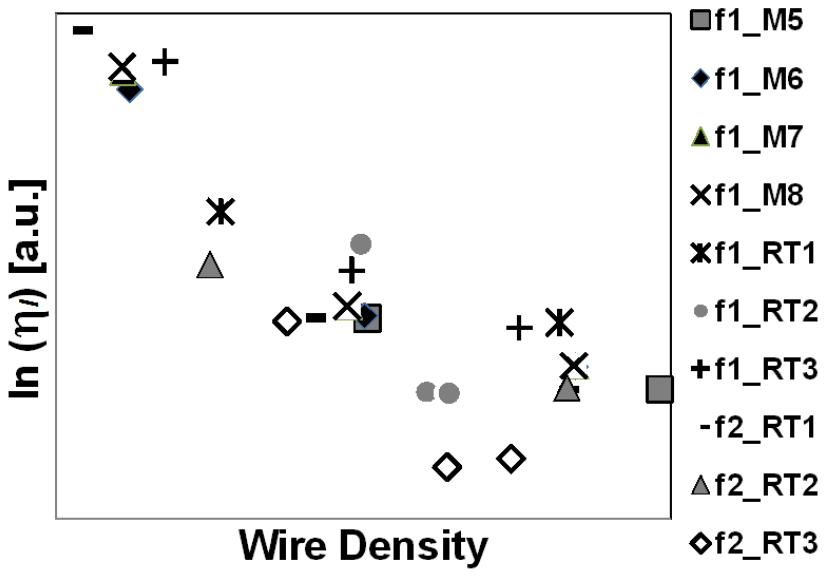


Figure 4.11: Wire density for Metal4–Metal6 and their characteristic lifetimes according to the $\sqrt{E}$ model.

4.8.1.5 Linewidth and Reliability

Linewidth impacts TF by affecting etching and photolithography [64]. As shown by the previous results η increases as linewidth increases, for a given linespace, because of the interaction of physical design with etching and photolithography [64]. For the circuits used in this chapter, increase in η with linewidth can be attributed to increasing minimum linespace as going from Metal1 to higher layers.

4.8.1.6 Timing And Reliability

Timing optimization is achieved through buffer insertion, changing gate location, and gate sizing. In terms of interconnect, densely routed areas raise issues of coupling capacitance which are addressed by ripping-up and re-routing the nets. Wire sizing is another way to obtain timing performance although it was not used.

Buffer insertion results in an increase in total wirelength, resulting in a decrease in reliability, as apparent from results. Moreover, gate re-placement needs to be managed carefully from reliability's perspective because re-placing gates to a crowded region can result in higher electric fields. If gate sizing is used for timing optimization then the goals of timing align with those of reliability. Increasing gate size increases the degree of freedom for wire-to-pin connections and reliability and vice-versa. Rip-up and re-routing aim at reducing wiring congestion and coupling capacitance, factors critical to reliability.

Metal3 is the critical layer when determining reliability for circuits optimized for timing performance. No trend was observed between timing and reliability because timing optimization generally uses heuristic algorithms instead of deterministic algorithms.

The results showed a strong correlation between the coverage in a given layer by the critical linespace group and the TF. For instance, for circuits $f1\_RT1$, $f1\_RT2$, and $f1\_RT3$, 99% of the lines are separated by two linespacing groups, $120nm$ and

$310nm$. The lifetime is determined by the $120nm$ linespace group. Interestingly, for the circuits optimized for timing, in every layer 95% of the lines had less than 3 linespaces between them and out of these three more than 50% linespaces were from the critical linespace group.

4.8.1.7 Impact of Change of Linewidth via Design Rule on Lifetime

As shown in the last chapter, changes in linewidth result in changes in linespace leading to changes in backend TDDB behavior. To assess how ARDE impacts characteristic lifetime, three different instantiations of the example radix-2 pipelined FFT 8 chip were generated. The reference layout was labeled $L1X$. Layouts $L3X$ and $L4X$ have three and four times wider Metal3 linewidth, respectively, than the linewidth of Metal3 in $L1X$. Figure 4.12 shows that the change in linewidth increases the characteristic lifetime for all of the layers, as well as for the chip.

The increase in lifetime for all layers can be attributed to a change in the routing

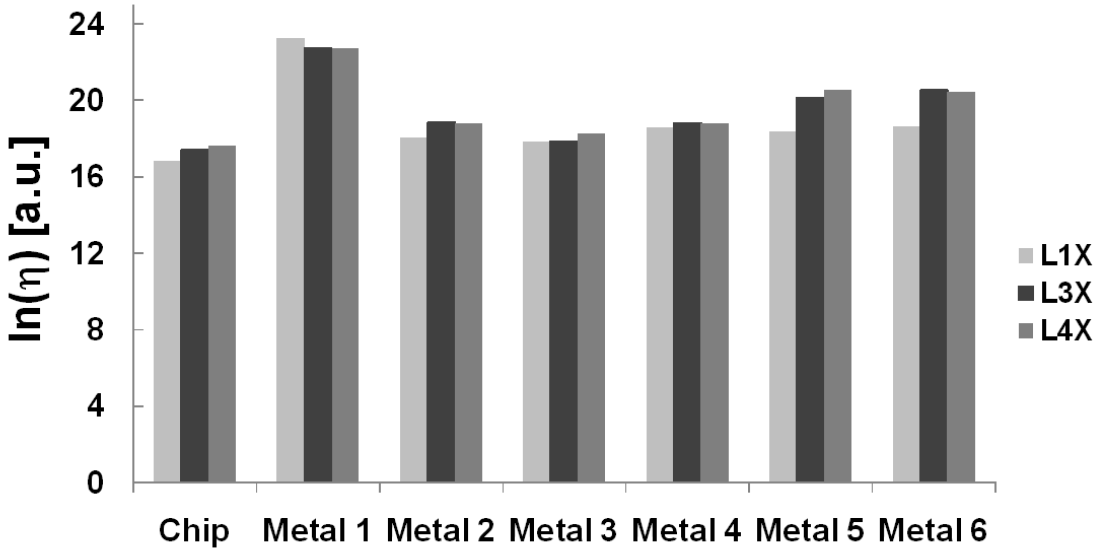


Figure 4.12: Characteristic lifetime for different instantiations of the FFT8 chip. The Metal3 linewidth of $L3X$ and $L4X$ is three and four times the linewidth of $L1X$, respectively.

for all layers due to the smaller number of routing tracks in Metal3. Hence, the increase in vulnerable area due to re-routing has a greater impact by degrading lifetime, outweighing any improvement in lifetime due to the use of wider Metal3 lines.

4.8.2 Results Based on Geometry and Function

Steady-state temperature of a point $p = (x, y, z)$ inside a thermal structure can be obtained by solving the heat equation

$$\nabla \cdot (k(p)\nabla T(p)) + S_{\mathrm{h}}(p) = 0, \tag{4.23}$$

where k is thermal conductivity, T is temperature, and S_{h} is volumetric heat source. This model can be implemented by meshing analyzed structure of the IC into thermal cells. To perform the thermal analysis, the layout of the IC is generated in DEF or GDSII format from Cadence SoC Encounter [154] and then static power analysis is performed, for a given circuit clock period and frequency (f) and logic cell switching activity (α_i), to determine power dissipation

$$P_i = \frac{1}{2}\alpha_i C_i V_{\mathrm{DD}}^2 f, \tag{4.24}$$

where C_i the loading capacitance of logic cell i and V_{DD} is the supply voltage. The layout along with the logic cell power dissipation is then used by the analyzer. The analyzer automatically generates the meshed structure for the IC along with the thermal conductivity and the volumetric heat source of each thermal cell. This information is used to perform thermal analysis using ANSYS FLUENT. Figure 4.13 shows the thermal map, with an activity factor of 0.5, for Metal3 of $f2\_RT3$.

4.8.2.1 Runtime of Thermal Simulations

The runtime of thermal analysis consists of the runtime for determining the percentage of material in each thermal cell to determine the thermal conductivity, the volumetric heat sources inside each thermal cell of the meshed structure, and the runtime for

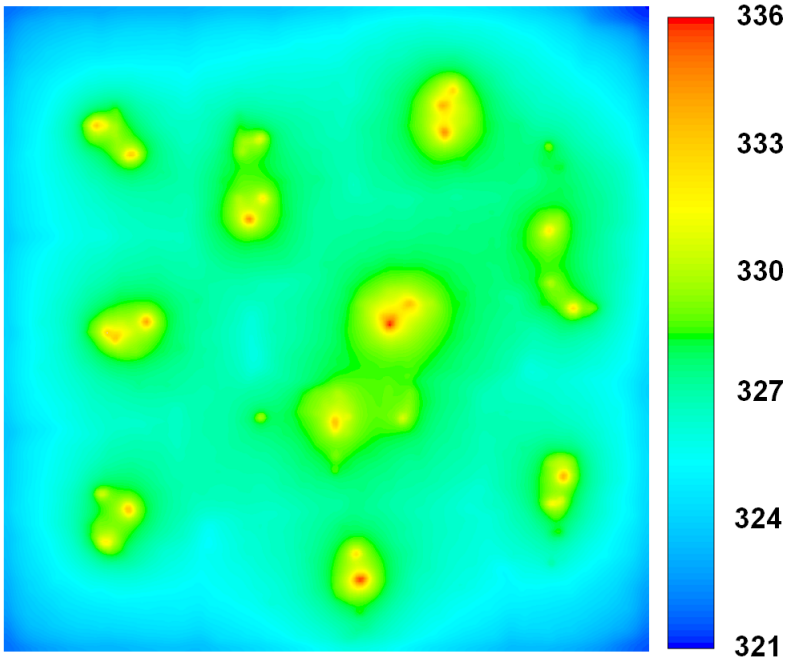


Figure 4.13: Thermal map of Metal3 of the circuit $f2\_RT3$ for an activity factor of 0.5.

solving the partial differential equations. The worst case complexity for the former is $O(n^2)$ and the average is $O(n \log n)$, n being the number of layout geometries. ANSYS FLUENT uses the finite volume method and its runtime varies between $1/40^{\text{th}}$ and $1/25^{\text{th}}$ of the time it takes to determine the percentage of the material in each thermal cell and the volumetric heat sources. Note that once the thermal analysis has been done, the layout statistics are generated, further information about the generation of layout statistics can be found in [156]. The runtime of layout statistics is given in the respective section. The layout statistics are integrated with the thermal profile of the chip.

4.8.2.2 Results

Figure 4.14 shows the results for the circuits simulated after incorporating their temperature profiles for a given signal activity level.

The trend among the models and the circuits remains the same after integrating the temperature profiles. Only the magnitudes of the characteristic lifetimes change.

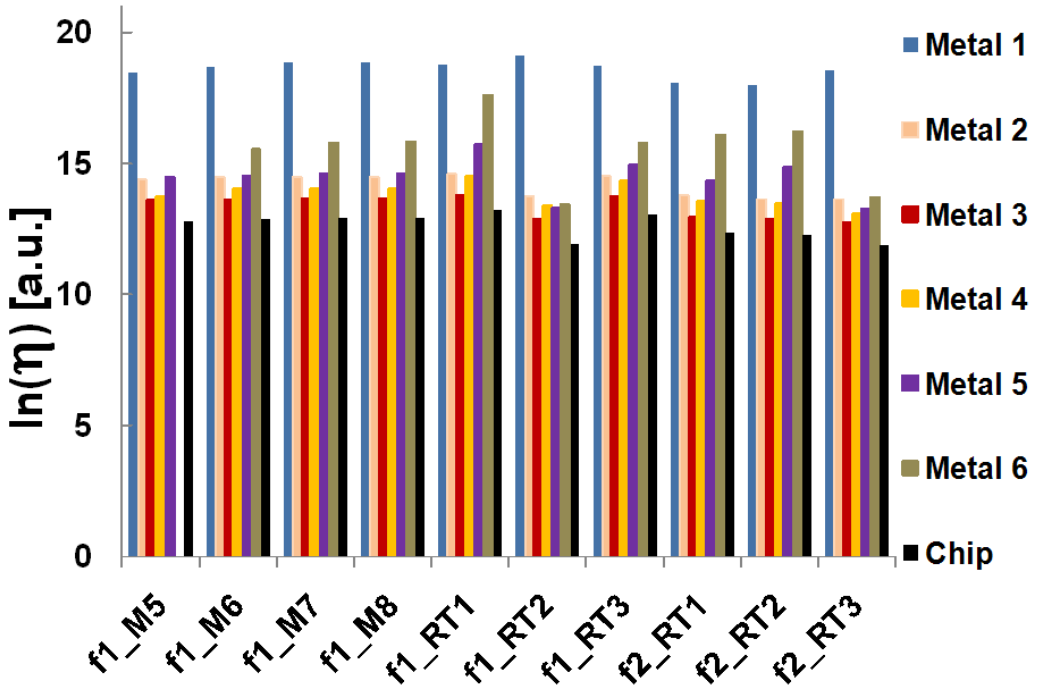


Figure 4.14: Characteristic lifetime for each layer and η for the chip for the circuits using the E Model and temperature profiles.

Figure 4.15 shows lifetimes with and without temperature profiles for the circuits for both the E model and the $\sqrt{E}$ model. Figure 4.16 shows the characteristic lifetime for all layers of $f1\_M5$ for both the E model and the $\sqrt{E}$ model, with and without the temperature map.

4.8.3 Temperature Map

Integrating the temperature profile in the methodology takes into account the variation in characteristic lifetime caused by the variation in on-chip temperature. However, for large layouts the best-case complexity for generating the thermal map is greater than the worst-case complexity for generating the layout statistics. Moreover, different input vectors affect the thermal map differently, thus requiring exhaustive thermal profiling, unless there formal methods to generate thermal profiles are used. Hence the efficacy of including thermal maps ultimately depends on the intended use of the simulator. If the simulator is being used for accurate reliability estimates, then

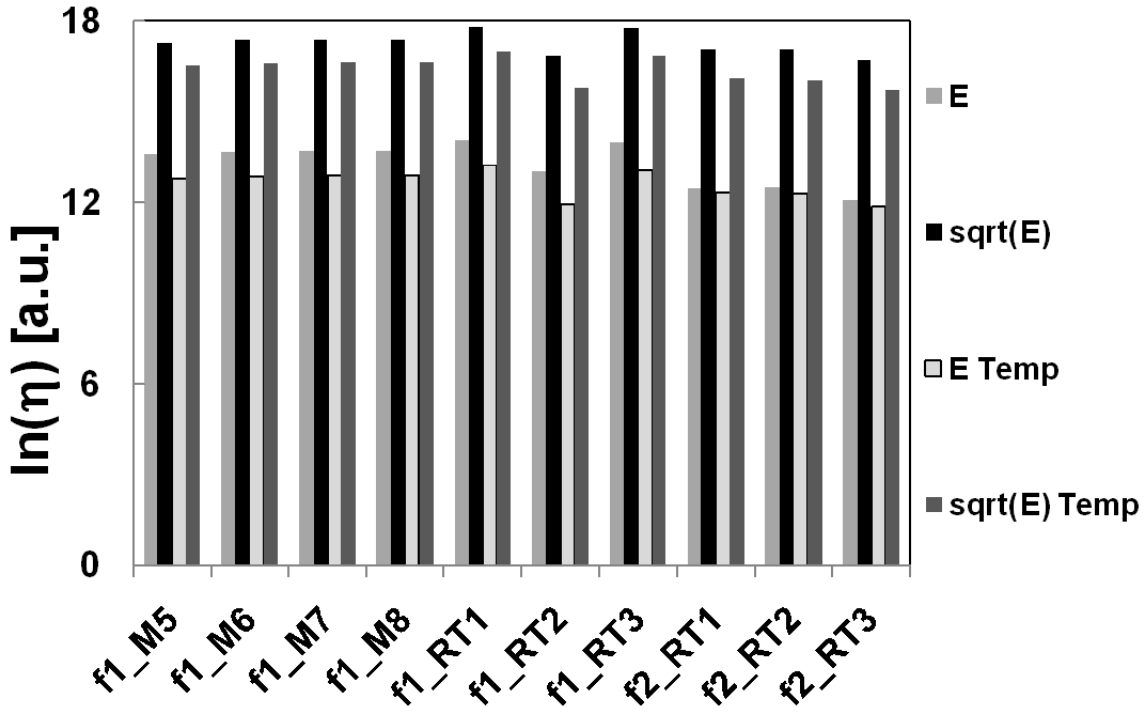


Figure 4.15: Characteristic lifetimes for the circuits according to the E model and the $\sqrt{E}$ model with and without temperature profile.

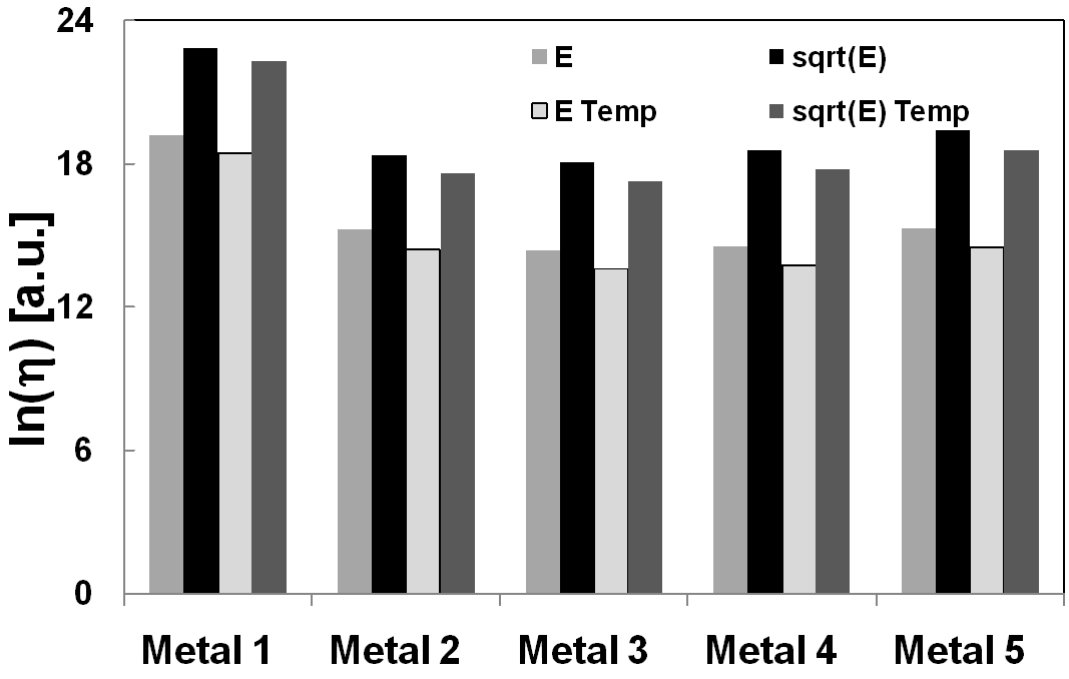


Figure 4.16: Characteristic lifetime for all layers of $f1\_M5$ with and without temperature according to the E Model and the $\sqrt{E}$ model.

thermal maps must be integrated. The same will be the case if the intention is to observe the effect of a particular class of input signals. However, if the intended use for the designer is to get some quick reliability numbers, then results without thermal maps can give somewhat of an accurate guess at best, or describe the range of lifetimes at worst.

4.9 Conclusions

A methodology was proposed to assess backend TDDB chip reliability. The methodology was developed in a way that other failure mechanisms can be integrated into it. Results from the simulator, built upon the proposed methodology, showed the feasibility of the approach. In doing so, the effect of layout on backend TDDB reliability was analyzed. Results showed an absence of any correlation between timing performance and reliability. Results also showed that greater wire coverage results in smaller lifetimes, as expected. It was demonstrated that the narrowest linespace group may not impact the chip lifetime critically. Instead, it is the linespace group with the highest coverage that is most instrumental in determining lifetime. It was also demonstrated that integrating temperature maps result in lower, though accurate, TF estimates.

The proposed methodology does not assume the design to be "as drawn" or that the failures are being caused by the layout features. Of course, if such were the case, then an extreme value distribution would not have been used. It was assumed that the layout is manufactured from the geometries in the test structures and modeling takes into account the invariance of Weibull statistics to area scaling, thus justifying extrapolations. Furthermore, it was assumed that any failure causing mechanisms that manifest themselves in test structures are reflected in the characteristic lifetimes used in the methodology.

PART II

VARIABILITY

CHAPTER V

DETERMINING THE IMPACT OF WITHIN-DIE VARIATION ON CIRCUIT TIMING

5.1 Introduction

Circuits manufactured with the same fabrication process flow exhibit variation in performance due to non-uniformity in process conditions. Timing analysis requires the characterization of cell libraries which provide delay and slew of an output signal as a function of input slew and output loading. The standard approach is to develop tables for delay and slew of the output for a variety of corner conditions, involving a combination of nominal, fast, and slow n- and p-channel devices.

Let's consider the delay function

$$d = f(p_i), \tag{5.1}$$

where p_i are process parameters. To find delay, data on die-to-die variation is collected on key parameters using the scribe line test structures. Variation in these parameters is assumed to be 100% correlated within a die. Hence, a sample of parameters can be collected from a population of various sites within a wafer and for a collection of lots. Sampling variation within a die is not required. Principal component analysis (PCA) [157] of the data provides a set of key independent parameters and corners, as in [158]. For digital circuits, these corners are associated with fast and slow n- and p-channel devices and are labeled as fast and slow corners. Worst case delay is the maximum delay among the corners. Let's suppose that the worst case corner parameter values are p_i^{WC}. The worst case delay is

$$d^{WC} = f(p_i^{WC}). \tag{5.2}$$

These models can provide accurate estimates of delay, provided that there is no within-die variation. At the worst case corner, within-die variability reduces a circuit's operating frequency and increases its power dissipation [159]. Since corner models do not account for within-die variation, except through the use of arbitrary guardbands on circuit performance specifications, circuits are vulnerable to unexpected yield loss due to within-die variability.

Incorporation of within-die variation requires that the models of cells include variations in output delay and slew as a function of process parameters. Without loss of generality, let's consider a linearized version of the delay function at the worst case corner for die-to-die variation:

$$d = d^{WC} + \sum_{i=1}^{n} c_i \Delta p_i + \epsilon, \tag{5.3}$$

where $\Delta p_i, i = 1, \ldots, n$, are changes in process parameters from nominal values, $c_i, i = 1, \ldots, n$, are sensitivities of delay to the process parameters, and $\epsilon = N(0, \sigma_{\text{intracell}})$ is a Normally distributed random variable with a standard deviation of $\sigma_{\text{intracell}}$, the standard deviation of the delay due to random uncorrelated variation within a cell. The purpose of this chapter is to model only the last two terms in Equation (5.3). In this analysis, purely random variation within a cell is separated, modeled with ϵ, from correlated variation, modeled as $\sum_{i=1}^{n} c_i \Delta p_i$. Hence, to construct such a model, it is important to consider the key sources of within-die variation in advanced technologies, and how they relate to the last two terms in Equation (5.3).

Within-die variation has both systematic and random components. Random variation is caused by statistical fluctuation in the number of dopant atoms per unit volume [160], which impacts the device threshold voltage (and is a function of device dimensions [79,112]), mobility variation, due to local mechanical strain [161,162] (which varies as a function of device length), and transistor channel length variation due to line edge roughness, induced by the polymer characteristics of photoresist [163,164]. All of these components are uncorrelated and are easily incorporated

in $\sigma_{\text{intracell}}$, except if there is spatial correlation. If there is spatial variation, then cells in closer proximity have a higher correlation. A number of papers have proposed methods to extract distance-based spatial correlations functions [165–168].

Systematic within-die variation is associated with design-process interactions, coming from process steps, such as lithography (proximity effect [169,170], lens aberrations [171, 172], flare [173, 174]), etching [63, 121, 125, 175] (microloading, aspect ratio dependencies), and chemical mechanical polishing [176–178] (erosion, dishing, etc.). These sources of variation impact transistor channel length, interconnect line width and height, and contact size, as a function of nearby geometries, position within a chip, and pattern density. Systematic variation may also come from details of chip design and operating conditions that can result in variations in temperature and supply voltage within a die. In this chapter, the focus is on process variation in minimum sized transistors only, since circuit speed is more sensitive to transistor variation than variation due to interconnect, even when delay is dominated by interconnect delays [179]. Hence, the main contributor to systematic within-die variation is lithography.

The range of interaction of lithography is the reticle, which may incorporate several die. Variation due to lithography is a function of pattern density (flare), location (lens aberrations), and neighborhood (the proximity effect). As a result, cell characteristics are a function of their placement, the local layout density, and the neighboring cells. In this chapter, data on within-die variation is looked at to determine the best methods to characterize process data to enable the selection of data points where simulations should be performed to estimate the impact of within-die variation.

This chapter is organized as follows. The next section considers prior work on process characterization for use in circuit simulation. Section 5.3 summarizes the extraction of systematic variation in channel length and presents a method to define a set of systematic corners using data and evaluates the effectiveness of these corners

in determining extreme circuit performances. Section 5.4 looks at random within-die variation and extracts spatial variation. Section 5.5 concludes the chapter.

5.2 Prior Work and Motivation

This chapter looks at random within-die variation with the aim of extracting spatial correlation functions for two major sources of variation: threshold voltage and channel length. Current voltage (I-V) curves are used to extract variation, where variation in drive current is attributed to not just variation in threshold voltage, as in [180], or channel length (CD), as in [159,181,182], but is a function of both threshold voltage and channel length. It is shown that, like in prior work, there is no spatial correlation in threshold voltage [106,183]. However, it is also shown that no spatial correlation is observed in channel length variation, after systematic variation is eliminated from the dataset.

Prior work has incorporated location-based dependence as a spatial variation function [181,184–186], where variation as a function of location is based on the statistics of device variation at each position in a grid that overlays the layout [181,184]. Similarly, location-based variation is extracted based on grid-based data. Neighborhood-based variation is also extracted, with principal components analysis, as a function of the distance of poly gates to nearby features, as in [170, 179]. The extracted neighborhood-based and location-based principal component models not only provide the mean shift in cell performance as a function of neighborhood and location, as considered in [187] for neighborhood-based variation, but also statistics on how this shift varies for a given process from die-to-die.

This chapter uses this information on the axes of systematic variation to find a minimum set of "systematic corners". The systematic corners define the points where statistical simulations must be performed to find timing extremes for within-die variation.

5.3 Systematic Channel-Length Variation

The range of systematic channel length (CD) variation is the reticle. Therefore, the appropriate test structure must involve measurements in different positions within the reticle, with different neighborhoods. A 25-point array of resistors is used, as in [172, 179], to extract systematic CD variation. The resistor array covers the reticle field. The data is for a single 25-point array, where each site has multiple measurements of CD as a function of neighborhood. Specifically, each site has eight vertically oriented resistors that are used in this study. Hence, the analysis in this chapter is illustrative of a methodology to analyze data that is available from a fabrication line, while the exact numerical results will be different for each fabrication line. The data is from an older technology generation, $0.18\mu m$, but the results are equally applicable to recent technologies. In addition, any method to extract CD variation across a reticle field would be suitable for this analysis.

Systematic variation due to lithography impacts CD as a function of neighborhood, location, and density. Other sources of systematic variation are not considered. Specifically, variation across the wafer is not considered, since it does not interact with neighborhood, location, and density-based variation. Instead, across-wafer variation is considered in defining the worst-case corner for die-to-die variation.

The dataset can be used to extract only neighborhood and location-based variation, and hence the focus is on only these components. It should be noted that location-based variation (from lens aberrations) interacts with neighborhood-based variation (from the proximity effect and Coma) [172]. Therefore, a model of within-die variation based on process data should take into account these interactions when defining systematic process corners. This can be done by combining the neighborhood and location-based data into a single dataset prior to analyzing it by principal components. This was not done here, because of the small size of the dataset.

5.3.1 Principal Component Analysis

Process data is characterized with principal components analysis, in order to obtain a statistical model of variation. Before beginning analysis, since this work is only concerned with within-die variation, the average for each die needs to be subtracted. In all analysis that follows, it is assumed that the die average has been subtracted from the dataset.

Let's suppose that the dataset has m categories, representing variation in neighborhood, location, and/or pattern density. The CD dataset contains P instances of m-dimensional vectors CD. Principal components analysis begins by computing the $m \times m$-dimensional covariance matrix from the CD dataset. Next, the eigenvalues and eigenvectors of the covariance matrix are determined. The rows of the $m \times m$-dimensional principal components transformation matrix, PCM, are the normalized eigenvectors of the covariance matrix. To map any m-dimensional vector CD to and from the principal components domain, the following transformation equations are required

$$\begin{aligned} PCS &= PCM \times (CD - A), \\ CD &= A + PCMI \times PCS. \end{aligned} \tag{5.4}$$

A is a vector of the averages. PCS is a vector of principal component scores in the principal component domain. $PCMI$ is the transpose of PCM.

The insignificant PCSs are found through determining the eigenvalues of the covariance matrix. Dimensional reduction is achieved by setting coefficients of PCM that correspond to the eigenvectors associated with small and insignificant eigenvalues to zero.

The inverse of the PCM matrix, $PCMI$, is used to transform the PCSs back to the original CD domain. The significant PCSs weight the basis vectors, rows of $PCMI$, to regenerate CD patterns.

A statistical distribution of CDs generates a statistical distribution of PCSs. For P CD vectors in the dataset, P n-dimensional vectors PCS are generated in the principal components domain, where n is the number of significant principal components. Unlike CDs, the distributions of PCSs are independent. For each element of PCS, say PCS_i, the standard deviation, σ_i, is determined. Let PCS_{ip} be the i^{th} element of the p^{th} vector, PCS. Then,

$$\sigma_i^2 = \frac{1}{P-1}\sum_{p=1}^{P} PCS_{ip}^2. \tag{5.5}$$

If Normally distributed, most of the distribution falls within $-3\sigma_i \leq PCS_i \leq 3\sigma_i$.

If there are n significant PCSs, then the PCS domain is n-dimensional and has 2^n corners. Let Ω be a $n \times 2^n$-dimensional matrix, whose elements in the i^{th} row are $\pm\sigma_i$. The columns of Ω are all combinations of $\pm\sigma_i, i = 1, \ldots, n$. For example, for the case with $n = 3$,

$$\Omega = \begin{bmatrix} -\sigma_1 & \sigma_1 & -\sigma_1 & \sigma_1 & -\sigma_1 & \sigma_1 & -\sigma_1 & \sigma_1 \\ -\sigma_2 & -\sigma_2 & \sigma_2 & \sigma_2 & -\sigma_2 & -\sigma_2 & \sigma_2 & \sigma_2 \\ -\sigma_3 & -\sigma_3 & -\sigma_3 & -\sigma_3 & \sigma_3 & \sigma_3 & \sigma_3 & \sigma_3 \end{bmatrix}. \tag{5.6}$$

The set of corners of the PCS domain is the columns of 3Ω. In the CD domain, these 2^n corners are the 2^n columns of

$$CD = A + 3PCMI \times \Omega. \tag{5.7}$$

These correspond to extreme variation in CD, given that the principal component scores are within the n-dimensional cube: $-3\sigma_i \leq PCS_i \leq 3\sigma_i$.

This standard approach to principal components analysis assumes that there are many instances of CD for all neighborhoods, locations, and densities. If the dataset is small, as in the used examples, it needs to be partitioned, to extract the impact of neighborhood, location, and density separately. As a result, each measurement has three categories (neighborhood, location, and density) and Q duplicates. Suppose that each category is associated with an index, i, j, and k, and duplicates

are associated with the index $q = 1, \ldots, Q$. Then each CD for the dataset is labeled as CD_q^{ijk}. Averages for each feature type are determined: $A^{loc(i)}, i = 1, \ldots, I$, $A^{nbd(j)}, j = 1, \ldots, J$, and $A^{den(k)}, k = 1, \ldots, K$, and the overall average, A. To analyze location, for example, a new dataset is formed by subtracting the effect of neighborhood and density. In other words,

$$\widetilde{C}\widetilde{D}_q^{ijk} = CD_q^{ijk} - (A^{nbd(j)} - A) - (A^{den(k)} - A). \tag{5.8}$$

From this revised dataset, the I-dimensional vectors, $\widetilde{C}\widetilde{D}$, are formed and the corresponding covariance matrix is computed. $\widetilde{P}\widetilde{C}\widetilde{M}$ and $\widetilde{A}$ are computed from the revised dataset, to provide the corresponding principal components transformation equations, as in Equation (5.4).

If data is not available for all locations and neighborhoods, in each location, the available data needs to be used to find the shifts: $A^{loc(i)} - A$, $A^{nbd(j)} - A$, and $A^{den(k)} - A$, to subtract components from the dataset that are not being analyzed. It may be necessary to compute these shifts with regression. However, it is important to check for any confounding in the dataset to ensure accurate results.

5.3.2 Neighborhood-Based Variation

The proximity effect and Coma cause linewidth to vary as a function of neighboring features.

The proximity effect causes linewidths in dense areas to be different than linewidths in isolated areas, as well as line end shortening, and corner rounding. The proximity effect is caused by variations in light intensity during exposure of the photoresist, resulting from the presence of neighboring features. This intensity variation modifies the exposure of photoresist on gate edges, which in turn translates into systematic variation in gate CDs.

Coma is a lens aberration that also causes the CD to vary as a function of its neighborhood. Coma becomes severe when making use of resolution enhancement

techniques, such as phase shift masks (PSM) and off-axis illumination (OAI). Analyzing the impact of Coma requires that for a specific pattern, features to the left must be distinguished from features to the right.

5.3.2.1 Background

In prior work, [187] proposes the use of model-based aerial image simulation to incorporate the proximity effect in cell library characterization. The neighborhood of the cell in its final placement is accounted for by considering 81 different neighborhoods. Hence, a cell is characterized for each of these 81 possible neighborhoods.

However, [187] provides only a mean shift in cell performance as a function of neighborhood. It is also important to consider how performance varies for a given process from die-to-die. Hence, in this work, the focus is on finding the axes of systematic within-die variation.

5.3.2.2 Data Analysis:Case 1

The major causes of variation in the neighborhood based patterns are defocus and exposure dose control. It is well known that dense transistors become wider under defocus, while isolated transistors become narrower under defocus, as illustrated by Bossung plots (Figure 5.1). To determine the range of variation in neighborhood-based patterns, data on neighborhood-based patterns is analyzed by principal components analysis. Here, principal components were not used to identify a set of independent process parameters, but instead the statistical distribution of device sizes as a function of their neighborhoods was analyzed.

For the sake of illustration, let's consider the simple classification system which categorizes gates according to their two distances to the nearest poly feature to the right and left. The two distance categories are labeled "small" and "large", where the small distance corresponds to two poly lines separated by the minimum distance design rule and the large distance corresponds to a poly line with no other poly lines

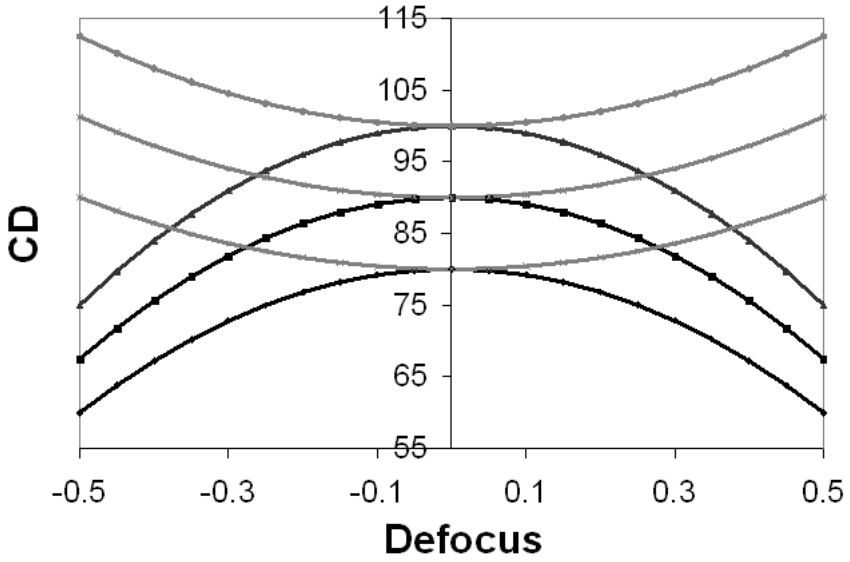


Figure 5.1: Example Bossung plot including variation due to exposure and defocus. The black curves correspond to isolated lines and the grey curves correspond to dense lines.

within the radius of interaction.

If there are two distance categories, small and large, and each transistor has two distances, to the left and to the right, neighborhood-based variation is characterized with a 4-dimensional vector, CD^{nbd}. To determine the CD for a specific transistor, the categories corresponding to each of the edges are determined, large or small, and the corresponding element in CD^{nbd} is looked up.

Other models of the proximity effect and Coma would also fit into the proposed modeling framework, where gates are categorized into discrete groups based on the characteristics of their neighborhoods. Neighborhood characteristics could be based on more than just the nearest feature, since the radius of influence for $193nm$ lithography is $400nm$. This means that the CD of a gate is influenced by many nearby features. One way to classify gates is through lithography simulation. The changes in CD induced by the aerial image could be used to categorize gates, since gates in each category are likely to undergo similar correction by the model-based optical

proximity correction algorithm.

As a simple illustration, let's consider the Bossung plot in Figure 5.1. Analysis of data from this figure results in two principal components. The data points are converted to the principal component domain to compute σ_1 and σ_2 which correspond to each of the two principal components. This is used to form the 2×4-dimensional matrix, Ω. Ω is converted to the CD domain, and the resulting four corners (columns of $CD = A + 3PCMI \times \Omega$) are shown in Figure 5.2, together with the original dataset of CDs for isolated and dense lines in Figure 5.1. Note that one principal component varies along the direction of exposure variation and the other varies along the direction of defocus variation.

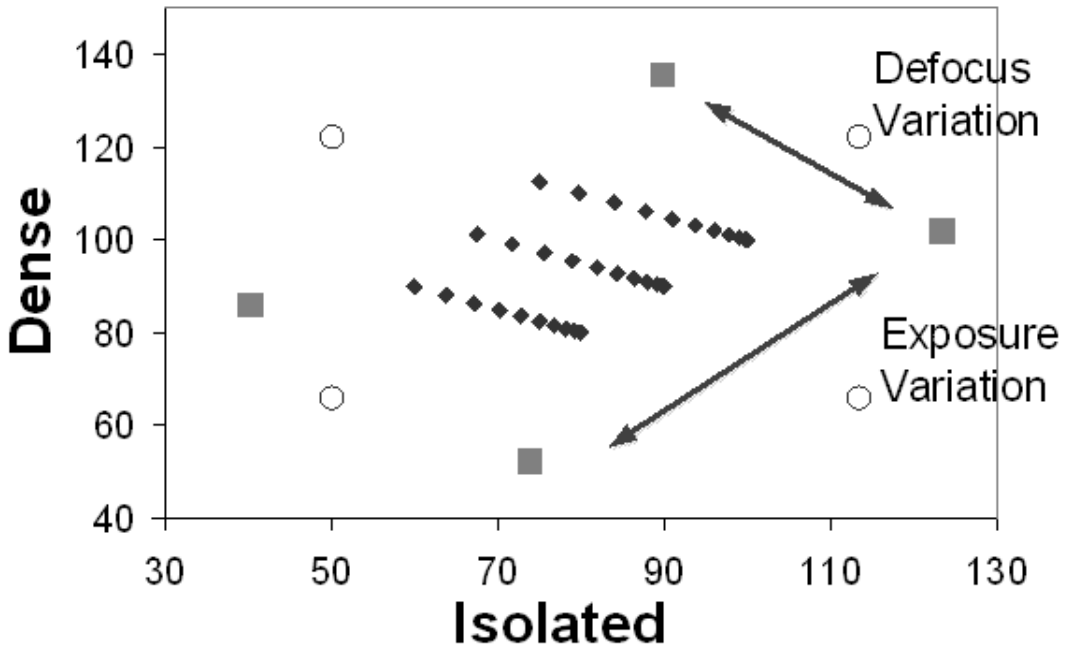


Figure 5.2: Isolated vs. dense CDs in Figure 5.1, together with the corners computed by principal components analysis. The data points are black, and the corners are grey. The three sigma corners computed without principal components analysis are shown as the open circles.

Figure 5.2 also shows the three sigma corners computed without principal components analysis. Under defocus and exposure variation, $CD_{\text{iso}} \in \left[CD_{\text{iso}}^{min}, CD_{\text{iso}}^{max}\right]$ and $CD_{\text{dense}} \in \left[CD_{\text{dense}}^{min}, CD_{\text{dense}}^{max}\right]$. In prior work [187], it was suggested that variation could be accounted for by finding the worst case, which is

$$\left(CD_{\text{iso}}, CD_{\text{dense}}\right) = \left(CD_{\text{iso}}^{max}, CD_{\text{dense}}^{max}\right). \tag{5.9}$$

This corner corresponds to the case with high exposure and very little variation due to defocus.

In the analysis, the two corners from principal components analysis near this point in Figure 5.2 are

$$\begin{aligned} CD_{\text{iso}} &= A_{\text{iso}} + 3PCMI_{1,1}\sigma_1 + 3PCMI_{1,2}\sigma_2, \\ CD_{\text{dense}} &= A_{\text{dense}} + 3PCMI_{2,1}\sigma_1 + 3PCMI_{2,2}\sigma_2, \end{aligned} \tag{5.10}$$

and

$$\begin{aligned} CD_{\text{iso}} &= A_{\text{iso}} + 3PCMI_{1,1}\sigma_1 - 3PCMI_{1,2}\sigma_2, \\ CD_{\text{dense}} &= A_{\text{dense}} + 3PCMI_{2,1}\sigma_1 - 3PCMI_{2,2}\sigma_2, \end{aligned} \tag{5.11}$$

where A_{iso} and A_{dense} are the average CDs corresponding to isolated and dense transistors, respectively, and $PCMI_{i,j}$ is the (i,j) entry in the $PCMI$ matrix. These two corners correspond to high exposure, but cover both the case with no defocus variation and the case with mismatch between isolated and dense transistors due to defocus. Hence, this approach covers a wider range of realistic process variations.

5.3.2.3 Data Analysis:Case 2

A small 4-dimensional dataset from a fabrication line was analyzed. The dataset contained variation due to location and neighborhood. Variation due to location was subtracted from the dataset. The raw data are shown in Figures 5.3 and 5.4.

From Figure 5.3 it can be seen that most of the variation for the isolated and dense data is in the direction of defocus. Figure 5.4 shows the data for the CDs that are dense on only one side.

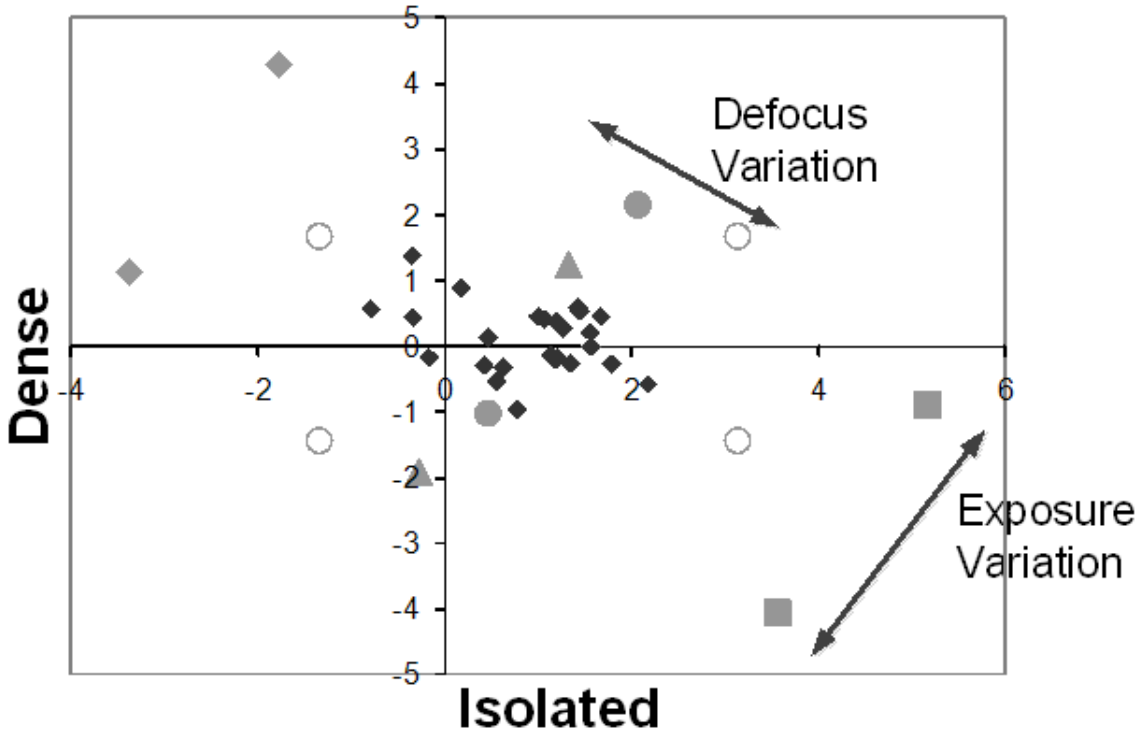


Figure 5.3: Isolated vs. dense CD data for a manufacturing dataset (in percent). The data points are black, and the corners are grey. The three sigma corners computed without principal components analysis are shown as the open circles.

Three principal components are needed to cover 100% of the variation. This results in eight corners. The projections of the corners are shown in Figures 5.3 and 5.4. Four corners represent variation due to dense/iso bias, and four corners represent variation due to Coma. Figures 5.3 and 5.4 also show the corners that would be computed without principal components analysis. These figures indicate that without principal components analysis, some of the corners correspond to unrealistic extremes that are not in the dataset.

The point (open circle) in the upper right quadrant in Figures 5.3 and 5.4 is the one that is suggested in [187] to determine the worst case timing under neighborhood-based variation. In contrast, this work suggests determination of the systematic corners that cover the actual systematic variation in the data.

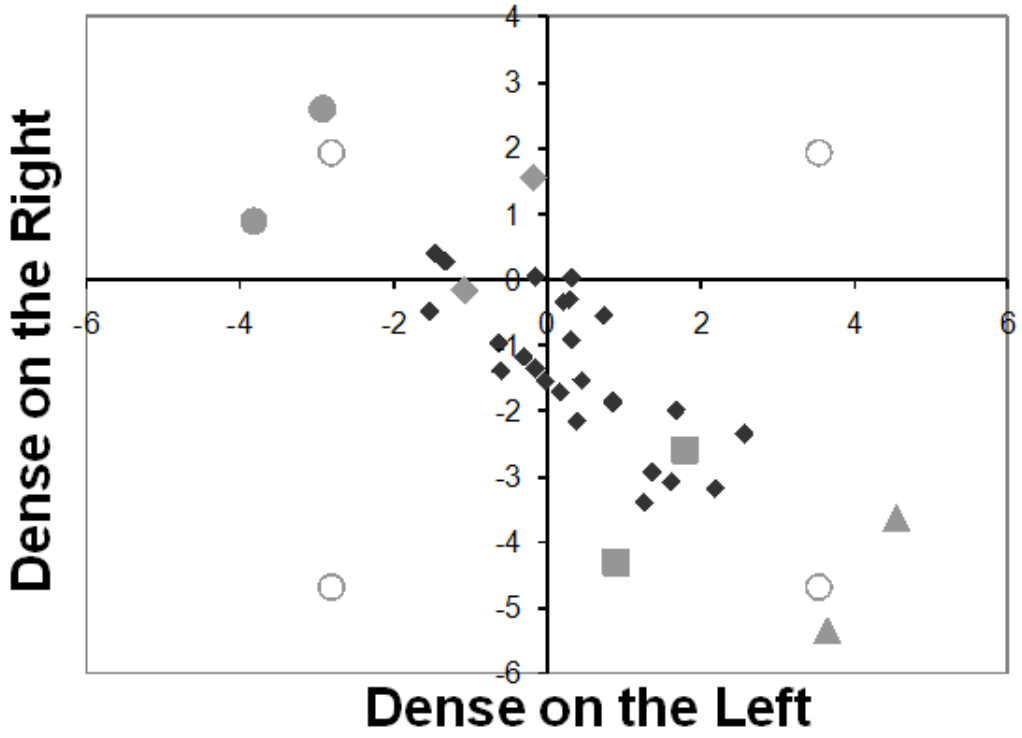


Figure 5.4: Manufacturing CD data for CDs that are only dense on the right vs. left (in percent). The data points are black, and the corners are grey. The three sigma corners computed without principal components analysis are shown as the open circles.

5.3.2.4 Impact on Circuit Performance Analysis

Let's consider the analysis of delay through a set of critical paths. If a path consists of g gates, then the delay, D, is

$$D = \sum_{i=1}^{g} d_i \ , \tag{5.12}$$

where d_i is the delay of the i^{th} gate in the path. Each of these delay terms, d_i, is a function of the CDs in the gates through variation of the load CDs and through variation of the drive current, which is also a function of the CDs in a gate. Hence, the delay of a path depends on all of the CDs in all of the gates involved. Let's suppose that there are g gates and a linear sensitivity of delay to CD. Then, the

relative variation in delay is

$$\frac{\Delta D}{D} = \sum_{i=1}^{g} \alpha_i \frac{\Delta CD_i}{CD_i} \ , \tag{5.13}$$

where α_i are constants. This equation indicates that the relative change in the delay of a path is a linear combination of the relative change in delay of the CDs of the gates in the path.

Some of the gates will be dense and some will be isolated. In fact, each of the gates will fall into one of the categories related to the proximity effect. Suppose that a classification system for the proximity effect has J categories, where $J = 2$ (iso and dense) for case 1, and $J = 4$ for case 2. For each category, sensitivities to variation based on neighborhood are the same. Therefore

$$\frac{\Delta D}{D} = \sum_{j=1}^{J} \alpha_j \frac{\Delta CD_j}{CD_j} \ , \tag{5.14}$$

where α_j are constants, relating the relative variation in the delay of a path to the relative variation in CD for a category.

The constants, α_j, are unique for each path. Moreover, total delay for a path is of the form:

$$D + \Delta D = \alpha'_0 + \sum_{j=1}^{J} \alpha'_j \frac{\Delta CD_j}{CD_j} \ , \tag{5.15}$$

where $\alpha'_j, j = 0, \ldots, J$, are constants. It should be noted that this formula applies equally well to all paths, even those heavily loaded with interconnect, with different ratios between the sizes of n-channel and p-channel devices, and with different input slew rates, provided that changes in CD are small.

Critical paths for within-die variation are paths that maximize Equation (5.15) for some process conditions. Under nominal conditions, the critical paths are those that maximize α'_0. Under variation due to exposure and defocus, other paths may become critical paths.

Finding a critical path for neighborhood-based systematic variation involves maximizing Equations (5.15) over a domain. The domain of interest is defined to be the

convex multidimensional cube, bounded by three sigma limits for each of the principal components. Hence, a linear function is being maximized over a convex domain. Under these conditions, the optimum occurs at a corner. Therefore, the critical path under variation can be determined by evaluating (5.15) at the corners, defined by Equation (5.7).

Let's consider the data described in Case 2 and a 64-bit pipelined multiplier circuit, Figure 5.5, that uses sequential logic and has 20 critical paths. Under nominal conditions, path 6 was found to be the longest. In order to find the maximum delay for all combinations of dense/iso bias, the eight corners were simulated (eight simulations). The corners identified two other critical paths, besides path 6. However, path 6 was found to be the longest when considering neighborhood-based variation.

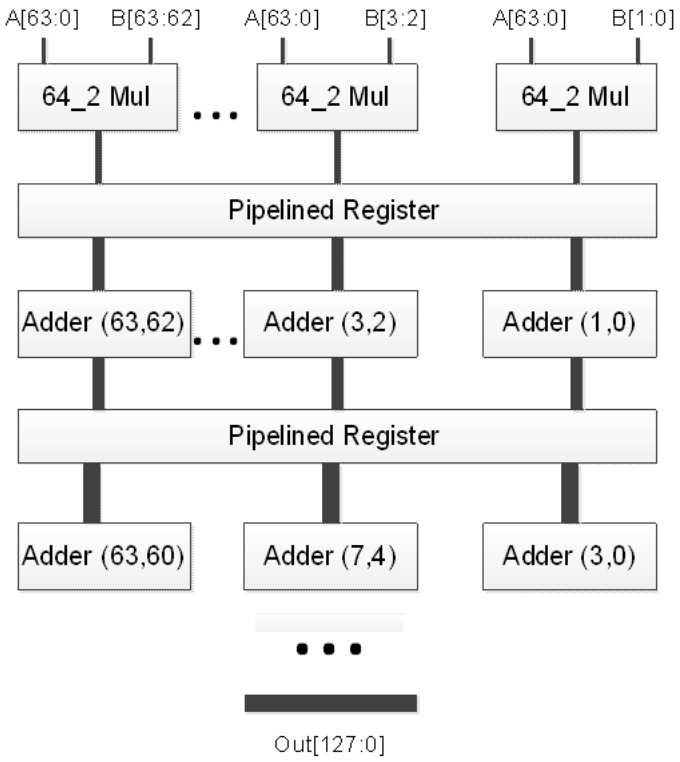


Figure 5.5: 64-bit pipelined multiplier using sequential logic. The circuit has 129 input pins and 129 output pins. It has 45453 gates and 45584 nets. It comprises of six pipelined stages and an area of $400\mu m \times 400\mu m$. The circuit is composed of 6 metal layers and has a target frequency of 667 MHz.

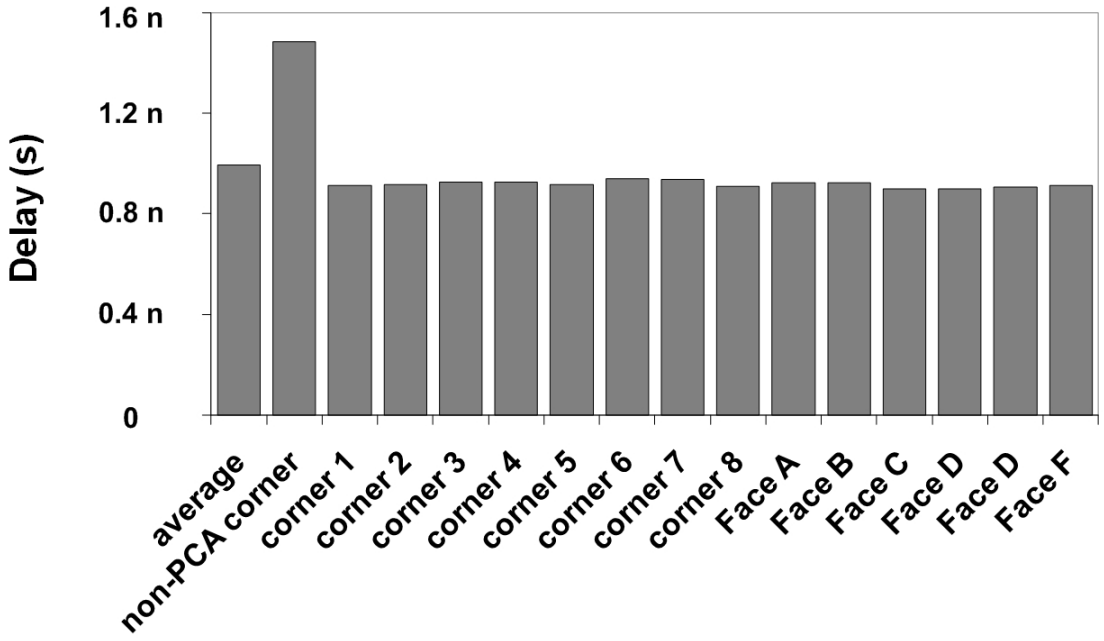


Figure 5.6: Worst case circuit delays at systematic corners found by principal components analysis of the data in Figures 5.3 and 5.4, in comparison with the worst case delay without principal components analysis. The figure also shows delays at the faces of the cube defining the principal components.

The difference between the proposed approach and the approach in [187] is the selection of the domain over which optimization is performed. The principal components domain more closely approximates the axes of variation in the data. As can be seen in Figure 5.4, without principal components analysis, corners can result in highly unlikely combinations.

Note that principal components analysis provides a much tighter bound on variation in comparison with a non-principal components analysis (open circles in Figure 5.3 and 5.4). The delays associated with each of the corners from principal components analysis are compared with the worst case delay without principal components analysis in Figure 5.6.

Analysis of the systematic corners may not determine all of the potential critical

paths because of nonlinearity in circuit responses. To check if there are likely to be any additional critical paths, six more cases were simulated on the faces of the cube defining the principal components domain. This analysis found the principal component corners can cover the delay. Also note that non-PCA corner shows the longest delay because the lengths in non-PCA corner are longer than nominal lengths resulting in smaller drive current.

Let's consider the data described in Case 2 and a set of paths whose nominal delay is exactly equal. Sensitivities to neighborhood-based variation will cause path delays to vary. 24 paths were considered. Under nominal conditions, path 15 was found to be the longest, because of dense/iso bias. However, die-to-die variation in the dense/iso bias causes variation in the critical paths. In order to find the maximum delay, the eight corners were simulated (eight simulations). The corners identified four other critical paths, besides path 15. Path 18 was found to be the longest when considering neighborhood-based variation.

The approach in [187] was also considered, which only requires one simulation at the $\pm 3\sigma$ point in each dimension. This method identified path 17 as the critical path, and the delay was found to be larger than the delay of path 15 at the worst case principal components corner. However, this point is beyond the $\pm 3\sigma$ bounding box in the principal component domain, as can be seen in Figures 5.3 and 5.4. In fact, it is in the direction of the 4^{th} principal component, which was eliminated, since it explained none of the variation in the dataset. And, it is 30σ away from the origin. Hence, analysis of variation without the use of principal components would correspond to introducing unrealistically large guardbands on circuit performance.

5.3.3 Location-Based Variation

Lenses have imperfections which are described by aberrations. Lens aberrations create optical path differences for each ray through the lens. Accounting for lens aberrations

involves determining the location of a transistor in the layout. Variability in transistor CD across the reticle could come from focus and exposure variation, lens heating, and lens characteristics.

This subsection compares a spatial correlation quad tree model with principal components analysis of the same data.

5.3.3.1 Data Analysis: Case 1

It is often observed that within-wafer variation is characterized with a donut and/or gradient profile due to the equipment used to process wafers. Most within-wafer variation can be handled as die-to-die variation. However, a component also impacts the within-die variation profile. To capture such a profile, a cross-wafer profile, shown in Figure 5.7, was generated. The wafer was assumed to include approximately 200 die. The data for all the sites was merged within the wafer and analyzed in two ways: with the quad tree model and with principal components.

The quad tree model is discussed in [181,188]. The hierarchical model is illustrated in Figure 5.8. The model contains three levels of hierarchy: level 0 through level 2. Level 0 represents correlation among all variables and level 2 represents uncorrelated variation. The dataset showed no variation associated with level 0, since this is die-to-die variation. Extraction of a quad tree model for the data involves finding four standard deviations for level 1 and 16 averages and standard deviations for level 2. Prior methodologies do not provide a clear method to extract these standard deviations. The model is extracted first by computing the 16 averages: $\mu_{2(1)}, \mu_{2(2)}, \mu_{2(3)}, \ldots$. Next, the covariances for level 1 are computed, i.e., the lower left sector of level 1 is computed based on the covariances between sectors 1, 2, 5, and 6. These are the level 1 variances, $\sigma^2_{1(1)}, \sigma^2_{1(2)}, \sigma^2_{1(3)}, \sigma^2_{1(4)}$. Each sector has six covariances, and the level 1 variance is the average. This level 1 variance is then subtracted from the variance in each sector to find the variance for each level 2 variable: $\sigma^2_{2(1)}, \sigma^2_{2(2)}, \sigma^2_{2(3)}, \ldots$.

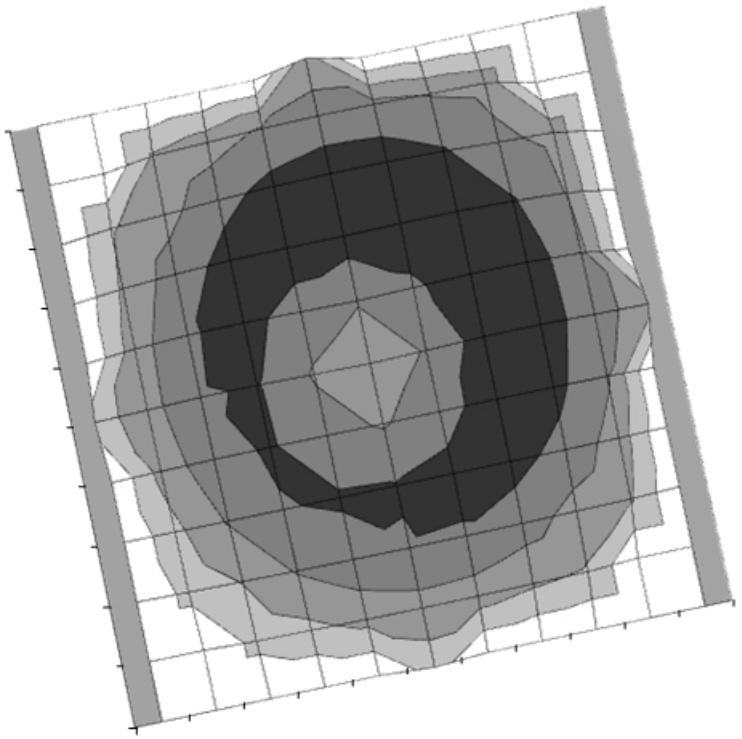


Figure 5.7: Cross-wafer donut pattern across a wafer with a slight gradient.

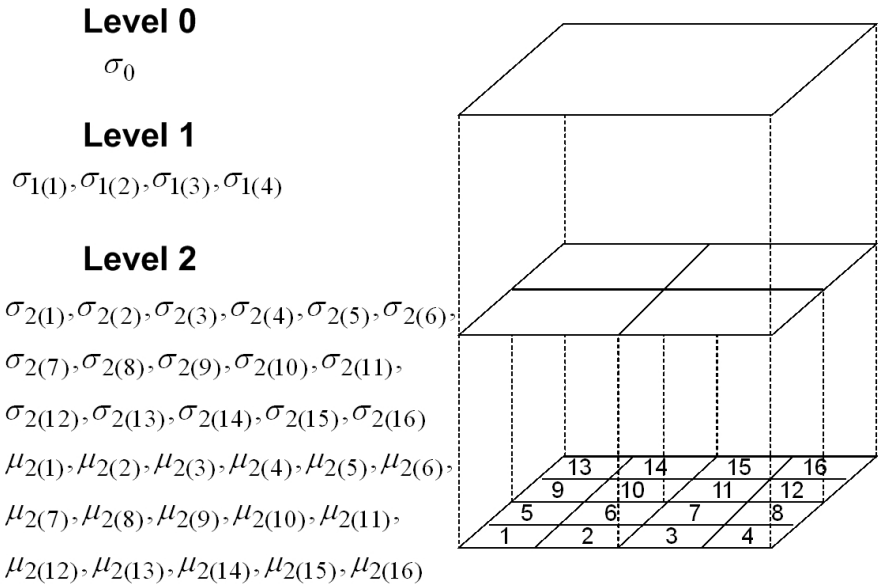
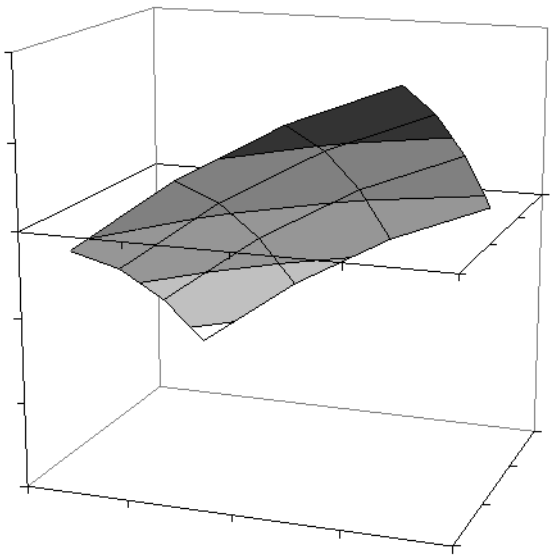
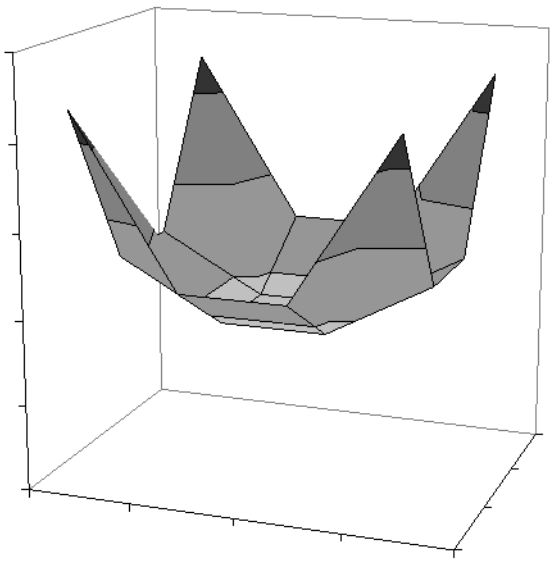


Figure 5.8: Quad-tree model and the parameters that are extracted at each level.

Figure 5.9 illustrates the extracted results, including the averages and variances extracted at each site. The level 1 variance is roughly equal in each sector and explains 41% of the variation in the data at the corners and 100% of the variation in the center of the die.



(a)



(b)

Figure 5.9: Extracted quad tree model, including (a) averages and (b) variances at each site. The average contour is the same for the principal component model.

To apply principal components analysis to the same dataset, each point in the grid is associated with a CD data point. All points on the grid are combined to form the 16-dimensional vector CD. The data for all die are combined to find the vector, A, and the matrices PCM and $PCMI$. The average vector A is illustrated in Figure 5.9(a) and is the same as computed with the quad tree model. The significant principal component basis vectors were determined. Two principal components explain 99.5% of the variation in the dataset. The extreme corners were found by mapping the original dataset to the principal component domain, and by analyzing the statistics of the PCSs to find σ_i for each of the two significant principal components. The corners of the space in the principal component domain characterize spatial variation extremes. They are converted back to the CD domain with Equation (5.7) and illustrated in Figure 5.10.

5.3.3.2 Data Analysis: Case 2

A manufacturing dataset is considered now using both the quad tree model and principal components analysis.

The data does not fit the classical quad tree model, since the data has an odd number of grid points. Therefore, a model with overlapping grids was implemented, as shown in Figure 5.11. This model extracts a distance-based correlation function.

This model contains several random variables, ranging from a variable representing correlation among all sites, to a set of variables representing correlation among adjacent sites. At the lowest level of the hierarchy, there are 25 random variables representing independent variation at each site, each with a unique mean and standard deviation. At the highest level of the hierarchy, there is one random variable representing correlation among all sites. In between, there are 16 random variables representing correlation between adjacent sites in each direction (level 3), nine random variables representing correlations between sites that are separated by one grid

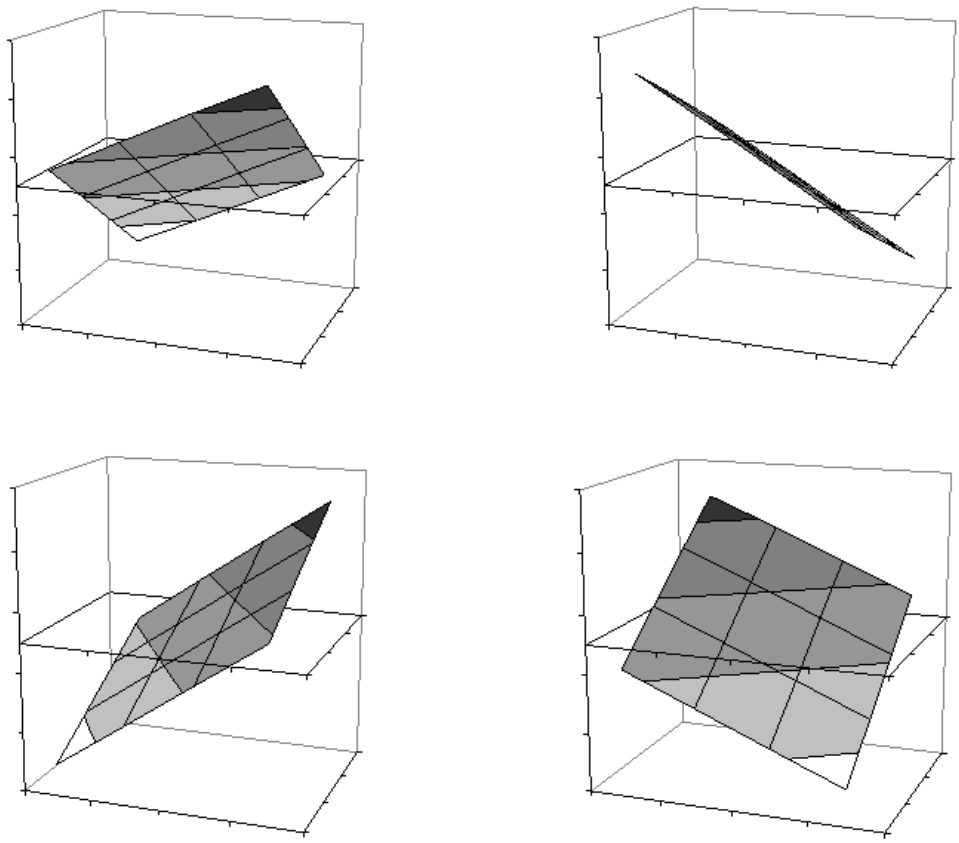


Figure 5.10: Location-based variation corners for within-die variation, determined by principal components of spatial variation of the donut contour in Figure 5.7, where the wafer contains approximately 200 die.

point (level 2), and four random variables representing correlations among sites that are separated by two grid points (level 1).

Die-to-die variation was eliminated in the dataset. Therefore, $\sigma_0 = 0$ by definition. The remaining standard deviations were found through least squares to optimize the fit to the data using the covariance between all sites in the dataset. Initially, it was assumed that all level 4 variables have a unique mean and standard deviation, and variables at other levels have the same standard deviation, i.e. σ_1, σ_2, and σ_3 were extracted. Only, the standard deviations in levels 3 and 4 were found to be non-zero. Therefore any spatial correlation is local correlation. In order to improve the fit, 16 level 3 standard deviations were computed to optimize the fit to the data. Hence, a

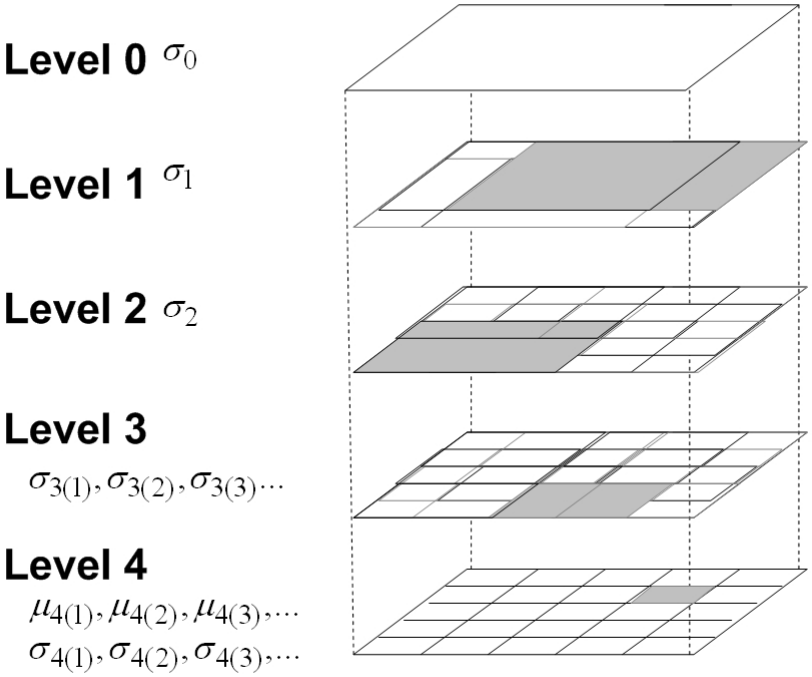


Figure 5.11: Quad tree model and the parameters that are extracted at each level, for a 5×5 grid.

total of 25 means and 41 standard deviations were extracted.

The quad tree model assumes positive covariances between all sites. As a result, the quad tree model has problems in fitting the data when there are negative covariances. When a negative variance was computed by least squares for a variable, the variable was eliminated from the dataset, and the standard deviations of the other variables were reoptimized.

Next, location-based variation was modeled using principal components analysis to determine spatial patterns. To apply principal components to spatial patterns, each point in the grid was associated with a CD data point. All points on the grid were combined to form the vector CD in Equation (5.4). Multiple instances of CD patterns were collected and analyzed to find the vector A, and the matrices, PCM and $PCMI$. Significant principal component basis vectors were determined.

The extreme corners were found by mapping the original dataset consisting of many instances of CD to the principal component domain, and by analyzing the statistics of the PCSs to find σ_i for each significant principal component. The corners of the space in the principal component domain characterize spatial variation extremes. They were converted back to the CD domain with Equations (5.7).

The spatial patterns depend on die size. Therefore, first, a large die where CD was measured on a 25-point grid was considered. It was found that for large die sizes (covering the entire reticle), four principal component cover over 90% of the variation in the dataset. Four principal components result in sixteen corners. Two of these process corners are illustrated in Figure 5.12.

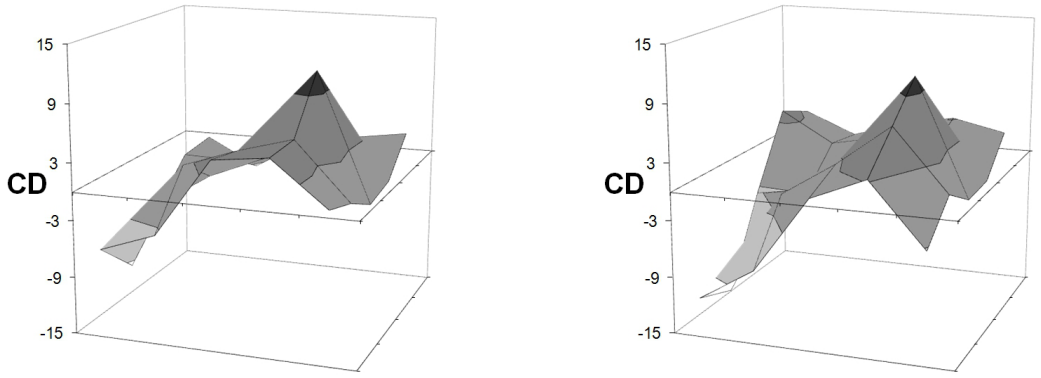


Figure 5.12: Two location-based corners, determined by principal components analysis of spatial variation using a 25-point grid covering the reticle field. CD variation is in percent.

The same data was partitioned into four nine-point grids and combined to analyze spatial variation for smaller die areas. It was found that over 90% of the variation was also explained by four principal components. This also results in 16 corners, two of which are illustrated in Figure 5.13.

It is likely that a more extensive dataset would result in a larger number of statistically significant corners. Also note that the full reticle principal component model contains a mean, four principal components, each represented by 25 numbers, and

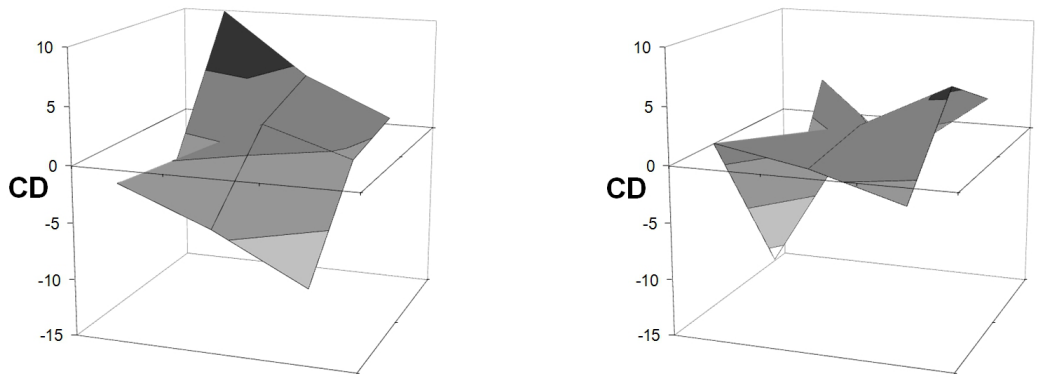


Figure 5.13: Two location-based corners, determined by principal components analysis of spatial variation for a smaller die size using a 9-point grid covering a fourth of the reticle field. CD variation is in percent.

four standard deviations associated with each of the principal components. Hence, the principal component model characterizes the data with 129 numbers, while the spatial correlation model characterizes the same data with only 66 numbers.

5.3.3.3 Impact on Circuit Performance Analysis

The relative change in the delay of a critical path is related to the relative change in CD, in accordance with Equation (5.14). Suppose that there are I sectors. Then the relative variation of delay due to CD variation is

$$\frac{\Delta D}{D} = \sum_{i=1}^{I} \alpha_i \frac{\Delta CD_i}{CD_i} \ . \tag{5.16}$$

However, unlike neighborhood-based variation, because of the large size of the reticle, critical paths fall within a single sector of the reticle. Consequently, for any path, all α_i are zero, except for the one coefficient corresponding to the sector of the path.

Let's suppose that a path is in sector i^*. Then, the total delay for a path is of the form:

$$D + \Delta D = \alpha'_0 + \alpha'_{i^*} \frac{\Delta CD_{i^*}}{CD_{i^*}} \ . \tag{5.17}$$

As with neighborhood-based variation, the paths that maximize α'_0 are the critical paths for the nominal process. Under variation, other paths may become critical,

especially those with large values of α'_{i*}.

Let's suppose that a large circuit is under consideration, where there are several critical paths in each sector, with the same value of α'_0. Let's also suppose that $\hat{\alpha}'_{i^*}$ is the maximum value of α'_{i*}, and there are several critical paths in each sector with $\alpha'_{i*} = \hat{\alpha}'_{i^*}$. Under these conditions, the candidate critical paths are those with maximum values of α'_0 and α'_{i*}. The process will select the manufactured critical paths from among these candidates.

Specifically, finding a critical path is equivalent to maximizing Equation (5.17) over a domain. The domain of interest is the convex multidimensional cube, bounded by three sigma limits for each of the principal components, such as those illustrated in Figures 5.12 and 5.13. Hence, a linear function is being maximized over a convex domain, and the optimum will occur at a corner. The critical path under variation can be determined by evaluating Equation (5.17) at the corners defined by Equation (5.7).

Let's first consider Case 1. The nominal worst case path would be found in sector $(4, 4)$, because of the average gradient across the die, illustrated in Figure 5.9(a). The critical paths at the corners are found to be in sectors $(4, 4)$, $(4, 1)$, $(1, 4)$, and $(1, 1)$, the four corners of the domain. The worst critical path is located in sector $(4, 4)$. The quad tree model associates larger standard deviations with the same corners. It would also find the worst case critical path in sector $(4, 4)$.

Figure 5.14 compares the variance at a few sites for the quad tree and principal components models. It can be seen that the two models extract the same variation at the same sites. The principal components model extracts a variety of gradients across the die, while the quad tree model attributes the uncertainty in the direction of the gradient to a larger standard deviation at the corners of the die. Simulations at the three sigma point for the quad tree model give the same result as simulations at the systematic corners from the principal components model.

Let's now consider Case 2 for the full reticle dataset. The nominal worst case

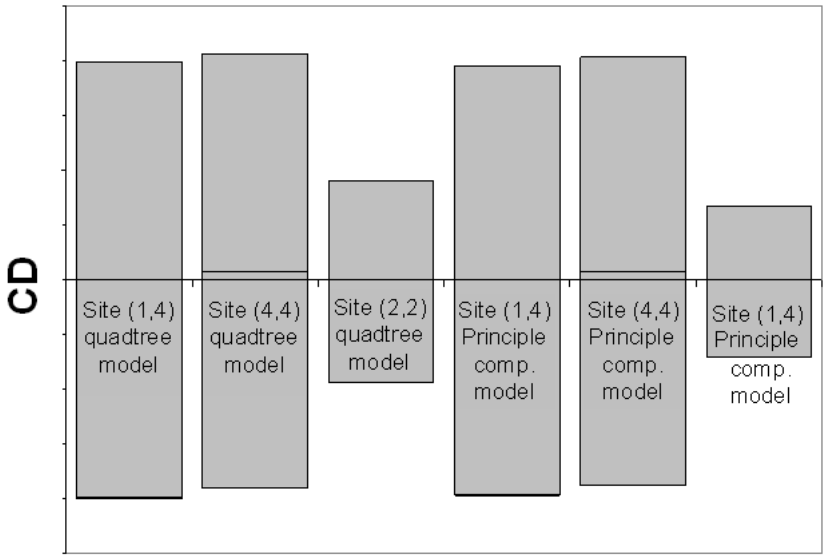
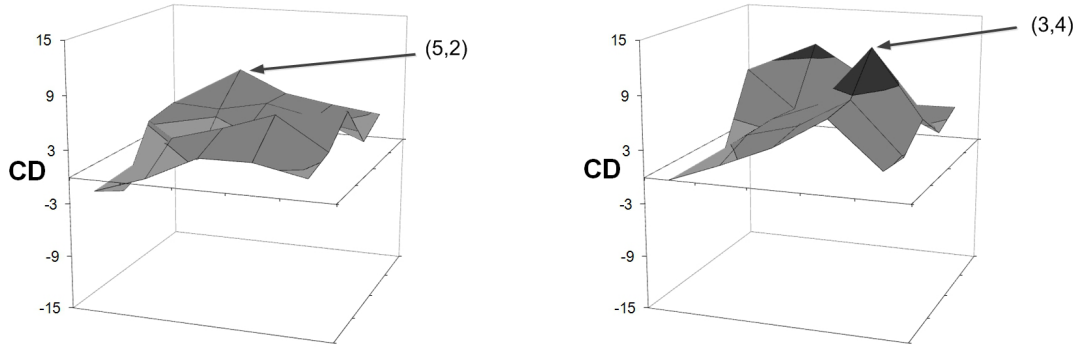


Figure 5.14: A comparison between $\pm 3\sigma$ variation for the quad tree spatial correlation model and the principal component model in three locations in the reticle for Case 1. The results at other reticle locations are similar.



(a) (b)

Figure 5.15: A comparison between $\pm 3\sigma$ variation for the (a) quad tree spatial correlation model and (b) principal component model in all locations in the reticle (in percent).

delay is in sector $(5, 2)$. The critical paths at the corners are found to be in sectors $(1, 4)$, $(2, 4)$, $(3, 3)$, $(3, 4)$, $(3, 5)$, $(4, 1)$, $(5, 2)$ and $(5, 3)$. The worst critical path is located in sector $(3, 4)$. The quad tree model, on the other hand, selects site $(5, 2)$ as the location of the worst critical path. The models do not match well. Figure 5.15 shows the maximum CD over all corners for the principal components model and the

maximum three sigma variation for the quad tree model in each sector.

The principal components model provides a larger bound on performance than the quad tree model. The reason for the difference is that the quad tree model does not fit the data well. This is because it cannot extract negative covariances between sectors. The dataset contains many negative covariances near the worst case $(3, 4)$ site, which are ignored in extracting the quad tree model.

5.4 Random Variation

In analyzing random variation, data must come from transistors that are closely spaced. To do this, data from a transistor array-based test structure, such as those in [189–191], implemented in $65nm$ technology was analyzed. Each of the transistors is individually addressable. The test structure contains $96,000$ SRAM-sized devices, placed in 1000 columns and 96 rows. The overall size of the array is $1250\mu m \times 110\mu m$. I-V characteristics were measured for 3840 of the identical devices. Variation in CD, ΔCD, and variation in V_{th}, ΔV_{th}, were extracted from the I-V characteristics.

The I-V curves are generated by applying voltages at the three terminals and measuring current. The gate and drain are accessed through the columns. The small number of rows ensures that the voltage drop along a column is small, i.e. less than $1mV$. The source connection is through the rows, and there is a parasitic IR drop at this source connection. Current is measured through a device between the transistor under test and ground. The row lines are contacted at both sides of the array, so that the IR drop along the row and the exact voltage at the source of the transistor under test can be computed. During all measurements, the terminals of the devices not being tested are driven negative to minimize leakage currents. The sum of currents from non-selected devices in any row was measured to be negligible.

5.4.1 Channel Length

A major source of randomness in channel length is line edge roughness, which is induced by the polymer characteristics of photoresist. In addition, corner rounding from the proximity effect, combined with overlay errors, can also result in random variation in channel length.

ΔCD was extracted first using the drain current, I_d, data when the transistor is in strong inversion, i.e. $V_{gs} = V_{dd}$ and $V_{ds} = 0.05V$. Under these conditions, I_d can be expressed as a function of terminal voltages [192]

$$I_d = \frac{W}{CD} \left(f\left(V_{gb}, V_{db}\right) - f\left(V_{gb}, V_{sb}\right) \right), \tag{5.18}$$

where V_{gb}, V_{db}, and V_{sb} are gate-body, drain-body, and source-body voltages, and W is the device width. At this bias, it can be shown that there is virtually no sensitivity of current to the threshold voltage. In addition, the device is in the linear region, and there is almost no contribution due to gate tunneling current, since at $V_{gs} = V_{dd}$, $I_d(V_{ds} = 0V)/I_d(V_{ds} = 0.05V) < 0.008$. The functions, $f(V_{gb}, V_{db})$ and $f(V_{gb}, V_{sb})$, have a small dependence on doping concentrations, which can be neglected when $V_{ds} \approx 0$.

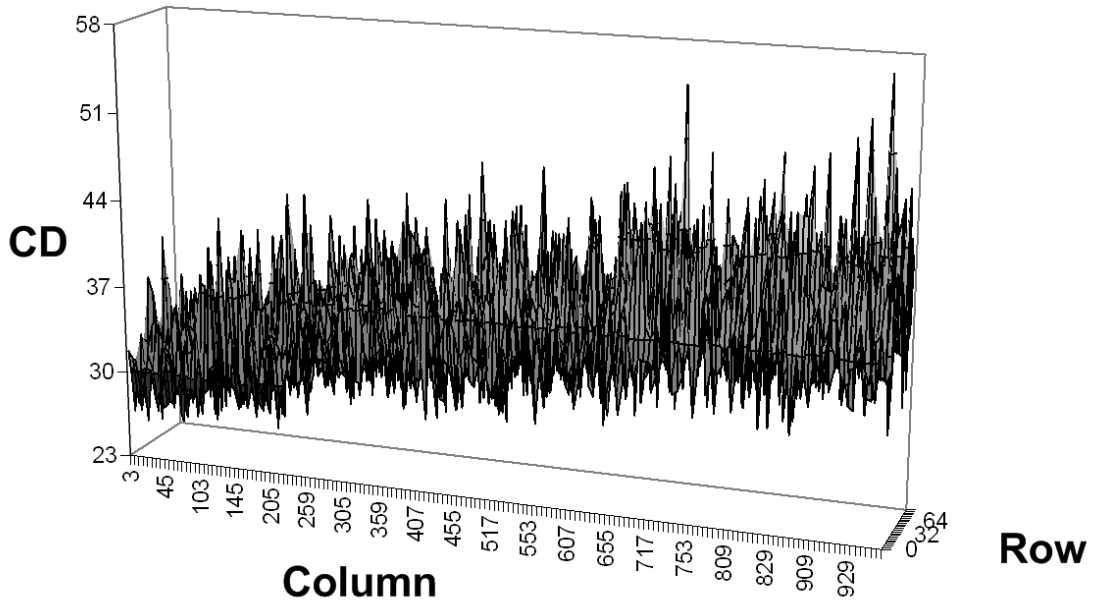
Let's then suppose that $I_d(V_{gs} = V_{dd}, V_{ds} \approx 0)$ and $\overline{CD}$ are averages for the entire array. Then

$$\frac{I_d\left(V_{gs} = V_{dd}, V_{ds} \approx 0\right)}{\overline{I}_d\left(V_{gs} = V_{dd}, V_{ds} \approx 0\right)} = \frac{\overline{CD}}{CD}\ , \tag{5.19}$$

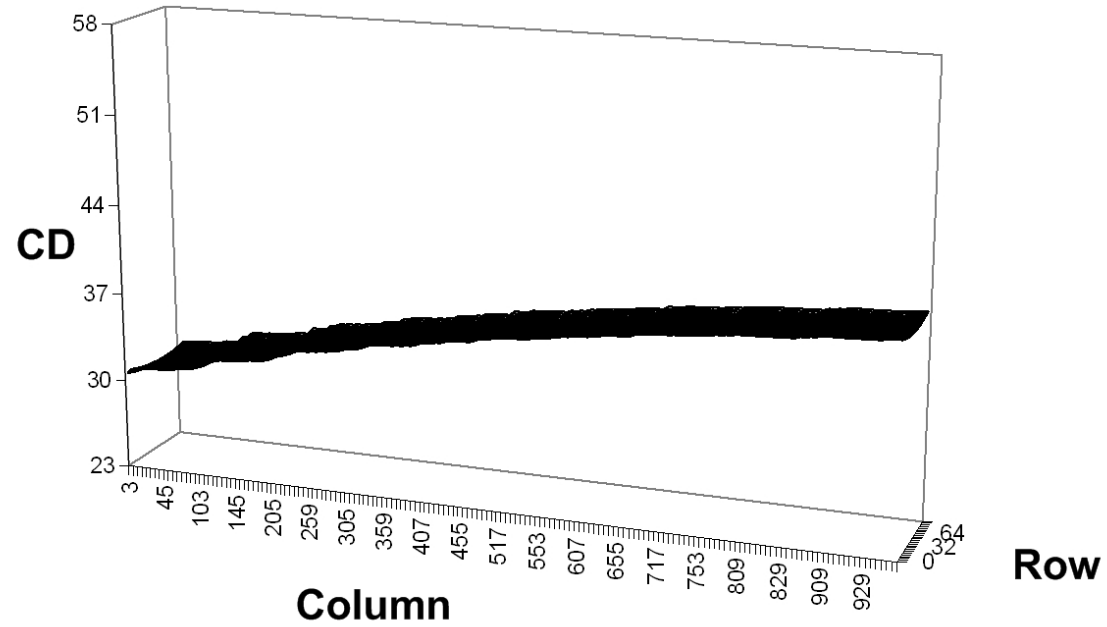
and

$$\begin{aligned} \Delta CD &= CD - \overline{CD} \\ &= \overline{CD} \left(\frac{I_d\left(V_{gs} = V_{dd}, V_{ds} \approx 0\right) - \overline{I}_d\left(V_{gs} = V_{dd}, V_{ds} \approx 0\right)}{I_d\left(V_{gs} = V_{dd}, V_{ds} \approx 0\right)} \right). \end{aligned} \tag{5.20}$$

To find ΔCD, $\overline{CD}$ must be known. Figure 5.16(a) shows raw data on channel length for a transistor array. The systematic spatial pattern extracted from the data is shown in Figure 5.16(b). In this dataset, the range of spatial variation is smaller



(a)



(b)

Figure 5.16: (a) Raw data on channel length for a transistor array. The horizontal scale describes the positions of the transistors in the array. (b) Model of the data indicating a trend as a function of position. The vertical scale of the two graphs is the same.

than the range of variation seen in Section 5.3.2 because the data in Section 5.3.2 covers the entire reticle field while this dataset only covers the transistor array. Hence, it is more desirable to extract spatial patterns from transistors placed throughout the reticle field to properly capture spatial variation.

The data displays almost no short distance spatial correlation between adjacent sites (< 0.01). The sample size for computing correlation was 1896. Hence, random within-die CD variation can be characterized by a single parameter, a standard deviation.

5.4.1.1 Model of CD Variation

When the CD models are combined together, an individual gate has a CD whose variation is determined by its nominal target CD, $\overline{CD}$, neighborhood, ΔCD^{nbd}, location, ΔCD^{loc}, and a purely random component, ΔCD^{r}:

$$CD = \overline{CD} + \Delta CD^{\text{nbd}} + \Delta CD^{\text{loc}} + \Delta CD^{\text{r}}. \tag{5.21}$$

The shifts due to neighborhood and location are modeled with principal components that explain shifts as a function of a mean and standard deviations of the significant principal component basis vectors:

$$\begin{aligned} \Delta CD^{\text{nbd}} &= A^{\text{nbd}} + PCMI^{\text{nbd}} \times \epsilon_{CD}^{\text{nbd}}, \\ \Delta CD^{\text{loc}} &= A^{\text{loc}} + PCMI^{\text{loc}} \times \epsilon_{CD}^{\text{loc}}, \\ \Delta CD^{\text{r}} &= \epsilon_{CD}^{\text{r}}. \end{aligned} \tag{5.22}$$

A^{nbd} and A^{loc} are neighborhood and location-specific mean shifts. $\epsilon_{CD}^{\text{nbd}}$ and $\epsilon_{CD}^{\text{loc}}$ are random variables describing die-to-die variations in neighborhood and location-based shifts in CD. ϵ_{CD}^{r} is a random variable accounting for random within-die variation in CD. The model requires the statistical characterization of $\epsilon_{CD}^{\text{nbd}}$, $\epsilon_{CD}^{\text{loc}}$, and ϵ_{CD}^{r}, to determine σ_{CD}^{nbd}, σ_{CD}^{loc}, and σ_{CD}^{r}, if $\epsilon_{CD}^{\text{nbd}}$, $\epsilon_{CD}^{\text{loc}}$, and ϵ_{CD}^{r} are normally distributed.

5.4.2 Threshold Voltage Variation

For small devices, the average number of dopant atoms in a channel is small. Random variation in the number of atoms gives rise to variation in threshold voltage. ΔV_{th} was extracted from the transistor array using subthreshold conduction, i.e., when $V_{gs} \approx 0$, $V_{ds} = V_{dd}$ [106, 193],

$$I_d = \frac{Wk}{CD} \exp\left(\frac{V_{gs} - V_{th} + \eta V_{ds}}{n\phi_t}\right)\left(1 - \exp\left(\frac{-V_{ds}}{\phi_t}\right)\right) , \tag{5.23}$$

where n in the subthreshold slope, i.e.,

$$n = \frac{1}{\phi_t}\frac{\partial\left(V_{gs}\right)}{\partial(\ln I_d)}, \tag{5.24}$$

k is a process-dependent constant, $\phi_t = {}^{kT}/_{q}$ is the thermal voltage, and η is the drain-induced barrier lowering (DIBL) coefficient. Then, if $\overline{I}_d(V_{gs} \approx 0, V_{ds} = V_{dd})$ and $\overline{CD}$ are averages for the entire array, we have that

$$\frac{I_d\left(V_{gs} \approx 0, V_{ds} = V_{dd}\right)}{\overline{I}_d\left(V_{gs} \approx 0, V_{ds} = V_{dd}\right)} = \frac{\overline{CD}}{CD} \exp\left(\frac{1}{\phi_t}\left(\frac{\overline{V}_{th}}{\overline{n}} - \frac{V_{th}}{n}\right)\right) \exp\left(\frac{V_{dd}}{\phi_t}\left(\frac{\eta}{n} - \frac{\overline{\eta}}{\overline{n}}\right)\right). \tag{5.25}$$

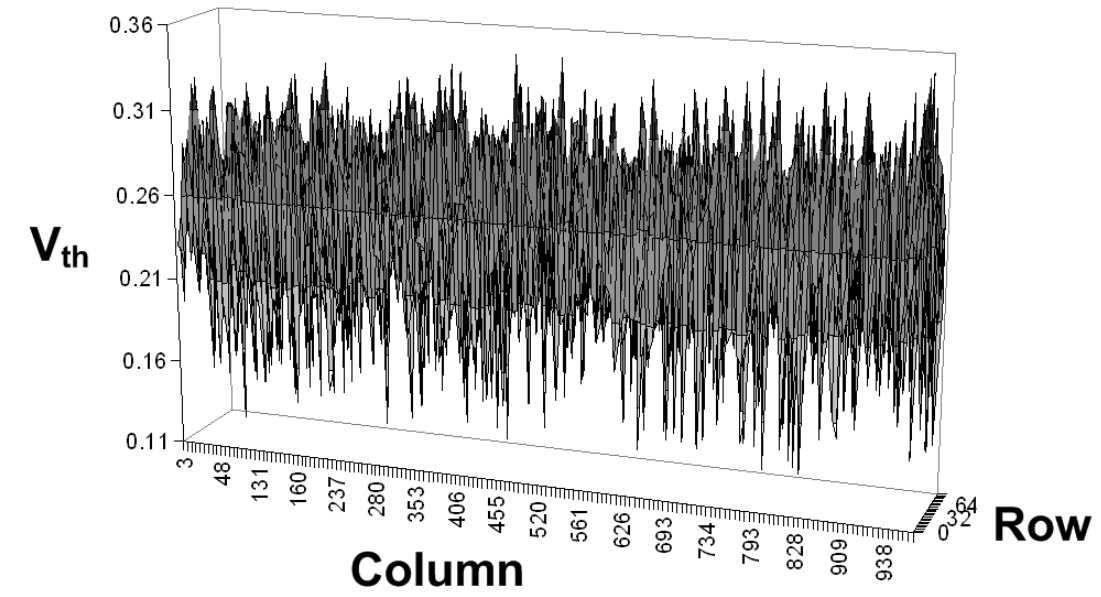
Therefore

$$\frac{V_{th}}{n} - \frac{\overline{V}_{th}}{\overline{n}} = V_{dd}\left(\frac{\eta}{n} - \frac{\overline{\eta}}{\overline{n}}\right) - \phi_t\left(\ln\left(\frac{I_d\left(V_{gs} \approx 0, V_{ds} = V_{dd}\right)}{\overline{I}_d\left(V_{gs} \approx 0, V_{ds} = V_{dd}\right)}\right) + \ln\left(\frac{CD}{\overline{CD}}\right)\right), \tag{5.26}$$

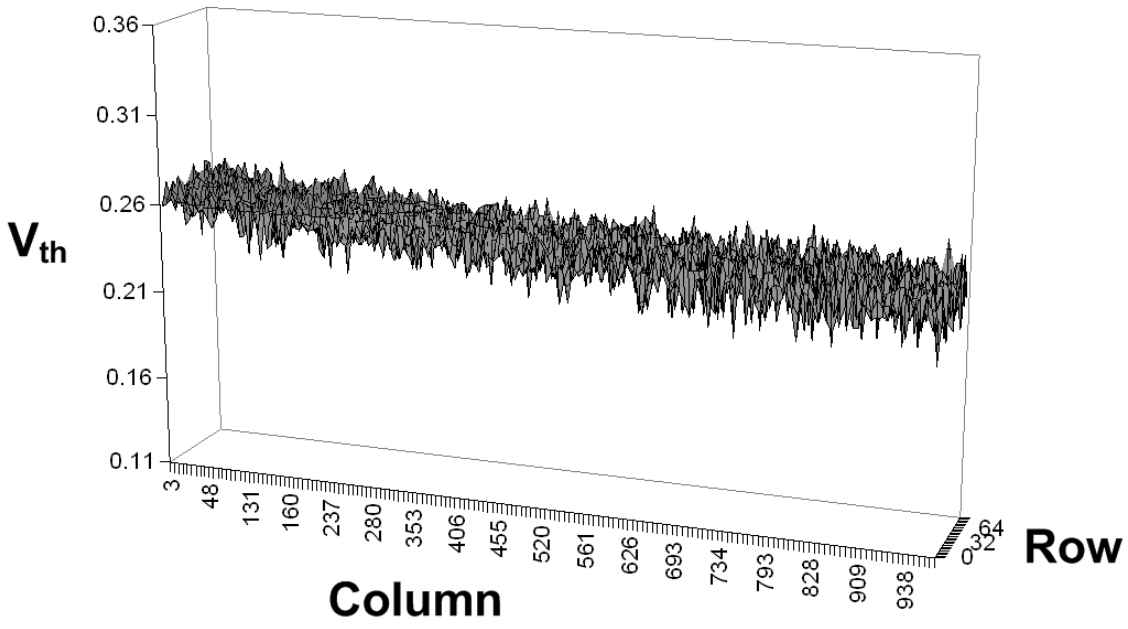
or equivalently, assuming small variations in the subthreshold slope, the DIBL coefficient, and the threshold voltage,

$$\begin{aligned}\Delta V_{th} = {} & V_{dd}\Delta\eta + \frac{\Delta n}{\overline{n}}\left(\overline{V}_{th} - \overline{\eta} V_{dd}\right) \\ & - \overline{n}\phi_t\left(\ln\left(\frac{I_d\left(V_{gs} \approx 0, V_{ds} = V_{dd}\right)}{\overline{I}_d\left(V_{gs} \approx 0, V_{ds} = V_{dd}\right)}\right) + \ln\left(\frac{CD}{\overline{CD}}\right)\right).\end{aligned} \tag{5.27}$$

In order to extract ΔV_{th} with Equation (5.27), the subthreshold slope is computed first using a rough estimate of V_{th} [194]. Next, $\overline{V}_{th}$ is computed using the constant current method [74] and data obtained by averaging the I-V curves for all sites. Finally, the DIBL coefficient is computed. This is done by finding $I_d(V_{gs} = \overline{V}_{th}, V_{ds} =$



(a)



(b)

Figure 5.17: (a) Raw data on threshold voltage for a transistor array. The horizontal scale describes the positions of the transistors in the array. (b) Model of the component of threshold voltage variation, explained by variation in channel length.

$0.1V) = I'_d$ and V'_{gs} such that $I_d(V'_{gs}, V_{ds} = V_{dd}) = I'_d$. The DIBL coefficient is computed as $\eta = (\overline{V}_{th} - V'_{gs})/(V_{dd} - 0.1V)$. The result is shown in Figure 5.17(a).

The threshold voltage is known to strongly depend on the channel length. This dependence is explicitly incorporated in Equation (5.27). However, other parameters in Equation (5.27) may display a dependence on channel length. Therefore, using the transistor array data, the variation in threshold voltage is modeled by

$$\Delta V_{th} = g\left(\Delta CD\right) + \epsilon^r_{V_{th}}, \tag{5.28}$$

where $g(\Delta CD)$ is a model relating ΔV_{th} to ΔCD, extracted using regression, and $\epsilon^r_{V_{th}}$ is the residual, not explained by variation in the channel length. Figure 5.17(b) shows the graph of the component of variation in threshold voltage that can be modeled as a function of channel length. It can be seen that a small fraction of variability in threshold voltage (7.5%) is explained by variation in channel length. This shows that the process dependent systematic component in the variation of threshold voltage is small.

The data display almost no correlation (< 0.02) between adjacent sites.

5.4.2.1 Model of V_{th} Variation

Neglecting the neighborhood and location component, the following is the model for the threshold voltage:

$$V_{th} = \overline{V}_{th} + g(\Delta CD) + \epsilon^r_{V_{th}}, \tag{5.29}$$

where $\epsilon^r_{V_{th}}$ is a random variable denoting within-die variation in the threshold voltage not explained by CD variation, which is characterized by $\sigma^r_{V_{th}}$ if $\epsilon^r_{V_{th}}$ is normally distributed.

5.5 Summary and Conclusion

Data has been analyzed to determine random and systematic components of within-die variation in channel length and threshold voltage.

Systematic variation is modeled with principal components. Principal components analysis defines the corners of the space of systematic variation. Moreover, if the domain of interest is defined by $\pm 3\sigma$ variation in the principal components, worst case delay due to within-die variation occurs at a corner of the systematic space. Hence, critical paths in the presence of systematic variability can be identified by simulating circuits at systematic corners.

Random variation was also analyzed. After elimination of spatial patterns in the dataset, it was found that there is virtually no spatial correlation in both channel length and threshold voltage between adjacent sites. Therefore, random variation can be modeled with independent random variables. Because of random variation, there may be several paths that are equally critical at each systematic corner. Hence, statistical static timing analysis [181,184–186,195–199] at systematic corners can find all such paths.

CHAPTER VI

CONCLUSION

6.1 Conclusions of the Research

6.1.1 Backend Low-k TDDB Reliability

This research looked at backend TDDB reliability of Cu/low-k interconnects from the view points of circuits and systems. Test structures have simple geometries, even if the physical features are similar to on-chip physical design. Chips, on the other hand, have a wide variety of physical design geometries on them. The goal of the reliability part of the thesis was to identify the sensitivities of TDDB to physical design geometries.

It was shown that backend low-k TDDB is consistent with Poisson area scaling, similar to gate dielectric TDDB. That is, for a given probability of failure, Cu/low-k interconnect systems with larger area will take a shorter time than an interconnect system with a comparatively smaller dielectric area. The research showed that despite the fact that low-k TDDB is a field driven mechanism, field enhancement at the tips of the interconnect has no impact on failure distributions. An analysis methodology was proposed, using the Poisson area scaling invariance of Weibull statistics, to separate the impact of area and electric field enhancement.

The research showed that failure data can be used to extract not only the distribution parameters but also the variation in the dimensions of the test structures. The latter is done by utilizing the failure distributions to determine the standard deviation of the die-to-die linewidth variation. Through the research, it was shown that while determining lifetime requirements for a given process not only the failure distributions parameters should be taken into account but the die-to-die linewidth variation

should also be considered. The research showed that large die-to-die linewidth variation adversely affects lifetime.

The research showed the impact of the process on low-k TDDB by identifying the interaction of etching with physical geometries. Specifically, it was shown that linewidth impacts TDDB lifetime. Narrower lines have shorter lifetimes because of the way narrower trenches are impacted by etching. Thus, caution must be exercised in all areas of process technology development and integration. The research showed that the linespace, which in turn determines the electric field and hence the TDDB reliability, is impacted by the dimensions of the neighboring Cu lines. It was shown that LER has a deleterious effect on TDDB performance, with the impact being greater for smaller linespaces. The research showed that pattern density does not impact TDDB reliability.

The research allowed the modeling of the impact of a wide variety of physical geometry features on backend low-k TDDB. These models and insights were then combined in a framework for a backend low-k TDDB chip reliability simulator for arbitrary chip layouts. The framework is modular in nature, thus other failure mechanisms can be included in the framework. The methodology for the simulator uses effective extrapolations of the stress data to chip while avoiding computationally expensive extensive physical simulations. The feasibility of the framework was demonstrated by the simulation results from different chip. Through the results of the simulator, the impact of different layout optimizations on bakcend low-k TDDB was observed.

6.1.2 CMOS Variability

Random and systematic variability in CMOS parameters, that manifests itself in device performance, was modeled. Systematic component of channel length variation, primarily caused by lithography, was modeled using principal component analysis. Random component of channel length variation was modeled by decoupling channel

length from other physical and electrical parameters. Similarly, the research showed that the random variation in threshold voltage can be modeled by decoupling it from other parameters. The research proposed new methods of extracting channel length and threshold voltage for a given process.

The research showed that the worst case delay due to within-die variation occurs at the corner of the systematic space. Thus, critical paths can be identified in the presence of systematic variability by simulating circuits at the systematic corners.

The research showed the lack of any spatial correlation in random variation in both the channel length and the threshold voltage. Thus, the research showed that random variation can be modeled with independent random variables. It was shown that, because of random variation, there may be several paths that are equally critical at each systematic corner. Hence, statistical static timing analysis at systematic corners can find all such paths.

The research showed that by modeling systematic components, that depend on the layout, with a finite set of corners, layout dependencies can be incorporated in the corner and the remaining variation does not depend on the layout.

6.2 Future Work

6.2.1 Backend Low-k TDDB Reliability

6.2.1.1 Testing with Small Linespaces

At smaller linespaces, additional effects may manifest themselves in failure data. Not only that, the failure distributions may show extrinsic failure modes. The effect of these on geometric sensitivity of backend low-k TDDB must be taken into account.

6.2.1.2 Irregular Geometries

On-chip physical geometries contain a wide variety of features. The features that were included in our simulator covered almost the entirety of the chips used for backend

TDDB reliability evaluation in this thesis. However, the impact of irregular geometries, that may be present in one chip and not in the other, on backend TDDB must also be analyzed. These include geometries that may enhance electric field between line ends and perpendicular lines, lines that may exacerbate fringing fields, and line ends that may be close but are at an awkward angle of each other. Similarly, the effect of optical proximity correction (OPC) on backend geometries and consequently on backend TDDB must be considered. If such geometries cause any vulnerability, then the vulnerable area must be considered, and hence included in the simulator, and during any attempt of extending test structure data to chips.

6.2.1.3 Interplay of TDDB with other BEOL failure mechanisms

Electromigration and stress induced voiding also impact interconnect reliability. The interplay of these two with backend low-k TDDB must be investigated. For instance, it is known that electromigration in Cu interconnects takes place along interfaces. The impact of voiding or hillocks on backend TDDB should be determined since Cu diffusion in the dielectric can be taken as a precursor of the eventual dielectric breakdown.

6.2.1.4 Chip Reliability

The proposed framework has the ability to incorporate other failure mechanisms in it. However, other than backend low-k TDDB and gate oxide breakdown no other failure mechanism follows a weakest link property. Thus, Weibull statistics can only be used for these mechanisms. The modularity of the proposed framework ensures that any other failure mechanism can be integrated into it, as long as it follows a weakest link property. It is desirable to have a chip reliability framework that considers all failure causing mechanisms. The results from the proposed framework are an evidence that the approach is a feasible solution. The incorporation of other failure mechanisms, modeled by different failure distributions, needs to be investigated from the view

point of integrating them in a similar way to the proposed methodology.

6.2.2 CMOS Variability

Understanding and modeling variation in electrical and physical parameters of CMOS drive current aids in circuit design by enabling the development of techniques to compensate for variation and to properly design in the presence of variation. While there are many models that attempt to do that, there is a need for a model that integrates different components of variation into a single framework. Thus variation must be characterized in a way that it can be used in the development of standard cell models. This will enable the incorporation of both random and systematic variation in circuit analysis. This calls for models of drive current that consider variation in physical and electrical parameters. Once such a model has been determined, models of delay can be developed. Similarly, such models of drive current and delay can incorporate environmental and temperature variations, leading to holistic models of variation in the operation of CMOS circuits and systems.

For some datasets, the number of systematic corners could be large. If there are too many systematic corners, the effort associated with simulating circuits at systematic corners could approach that needed for Monte Carlo analysis. To reduce the computational effort, algorithms should be developed to efficiently explore the space defined by the systematic corners, to avoid simulation of all systematic corners.

REFERENCES

[1] C. T. Sah, "Evolution of the MOS transistor - from conception to VLSI," *Proc. of the IEEE*, vol. 76, no. 10, pp. 1280–1326, 1988.

[2] R. H. Dennard, F. H. Gaensslen, V. L. Rideout *et al.*, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.

[3] R. H. Dennard, "Scaling limits of silicon VLSI technology," ser. Physics and Fabrication of Microstructures and Microdevices. Proceedings of the Winter School, M. Kelly and C. Weishbuch, Eds. Berlin, West Germany: Springer-Verlag, 1986, pp. 352–369.

[4] M. T. Bohr, "Interconnect scaling - the real limiter to high performance ULSI," in *Technical Digest, Int. Electron Devices Meeting (IEDM)*, 1995, pp. 241–244.

[5] R. Rosenberg, D. C. Edelstein, C. K. Hu *et al.*, "Copper metallization for high performance silicon technology," *Annual Review of Materials Science*, vol. 30, pp. 229–262, 2000.

[6] D. C. Edelstein, G. A. Sai-Halasz, and Y. J. Mii, "VLSI on chip interconnection performance simulations and measurements," *IBM Journal of Research and Development*, vol. 39, no. 4, pp. 383–401, 1995.

[7] P. H. Townsend, S. J. Martin, J. Godschalx *et al.*, "SiLK polymer coating with low dielectric constant and high thermal stability for ULSI interlayer dielectric," in *Proc. Materials Research Society*, vol. 476, 1997, pp. 9–17.

[8] S. J. Martin, J. P. Godschalx, M. E. Mills *et al.*, "Development of a low-dielectric-constant polymer for the fabrication of integrated circuit interconnect," *Advanced Materials*, vol. 12, no. 23, pp. 1769–1778, 2000.

[9] E. M. Zielinski, S. W. Russell, R. S. List *et al.*, "Damascene integration of copper and ultra-low-k Xerogel for high performance interconnects," in *Technical Digest, Int. Electron Devices Meeting (IEDM)*, 1997, pp. 936–938.

[10] *National Technology Roadmap for Semiconductors.* Semiconductor Industry Association, 1997.

[11] *International Technology Roadmap for Semiconductors.* Semiconductor Industry Association, 1999.

[12] W. W. Lee and P. S. Ho, "Low-dielectric-constant materials for ULSI interlayer-dielectric applications," *Materials Research Society Bulletin*, vol. 22, pp. 19–27, 1997.

[13] D. Pramanik and A. N. Saxena, "VLSI metallization using Aluminum and its alloys. I," *Solid state technology*, vol. 26, no. 1, pp. 127–133, 1983.

[14] ——, "VLSI metallization using Aluminum and its alloys. II," *Solid state technology*, vol. 26, no. 1, pp. 131–138, 1983.

[15] J. D. McBrayer, R. M. Swanson, and T. W. Sigmon, "Diffusion of metals in Silicon Dioxide," *Journal of the Electrochemical Society*, vol. 133, no. 6, pp. 1242–1246, 1986.

[16] S. S. Wong, A. L. S. Loke, J. T. Wetzel *et al.*, "Electrical reliability of Cu and low-k dielectric integration," in *Proc. Materials Research Society*, vol. 511, 1998, pp. 317–327.

[17] A. L. S. Loke, J. T. Wetzel, R. Changsup *et al.*, "Copper drift in low-k polymer dielectrics for ULSI metallization," in *Digest of Technical Papers Symp. on VLSI Technology*, 1998, pp. 26–27.

[18] W. Zhen-Cheng, W. Chau-Chiung, W. Ren-Guay *et al.*, "Electrical reliability issues of integrating thin Ta and TaN barriers with Cu and low-k dielectric," *Journal of the Electrochemical Society*, vol. 146, no. 11, pp. 4290–4297, 1999.

[19] R. Gonella, P. Motte, and J. Torres, "Time-dependent-dielectric breakdown used to assess copper contamination impact on inter-level dielectric reliability," in *Final Report Integrated Reliability Workshop (IRW)*, 2000, pp. 189–190.

[20] Z. C. Wu, Z. W. Shiung, C. C. Wang *et al.*, "Electrical reliability issues of integrating low-k dielectrics with Cu metallization," in *Proc. IEEE Int. Iterconnect Technology Conference (IITC)*, 2000, pp. 82–84.

[21] W. Zhen-Cheng, C. Chiu-Chih, W. Wei-Hao *et al.*, "Leakage mechanism in Cu damascene structure with methylsilane-doped low-k CVD oxide as intermetal dielectric," *IEEE Electron Device Letters*, vol. 22, no. 6, pp. 263–265, 2001.

[22] S. U. Kim, T. Cho, and P. S. Ho, "Leakage current degradation and carrier conduction mechanisms for Cu/BCB damascene process under bias-temperature stress," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1999, pp. 277–282.

[23] J. Noguchi, N. Ohashi, J. Yasuda *et al.*, "TDDB improvement in Cu metallization under bias stress," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2000, pp. 339–343.

[24] J. , T. Saito, N. Ohashi *et al.*, "Impact of low-k dielectrics and barrier metals on TDDB lifetime of Cu interconnects," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2001, pp. 355–359.

[25] A. M. Ionescu, G. Reimbold, and F. Mondon, "Current trends in the electrical characterization of low-k dielectrics," in *Proc. Int. Symp. on Quality Elec. Design (ISQED)*, 1999, pp. 27–36.

[26] W. Y. Shih, M. C. Chang, R. H. Havemann *et al.*, "Implication and solutions for Joule heating in high performance interconnects incorporating low-k dielectrics," in *Digest of Technical Papers Symp. on VLSI Technology*, 1997, pp. 83–84.

[27] G. W. Ray, "Low dielectric constant materials integration challenges," in *Proc. Materials Research Society*, vol. 511, 1998, pp. 199–211.

[28] A. Strong, E. Wu, R. Vollertsen *et al.*, *Reliability Wearout Mechanisms in Advanced CMOS Technologies*. Wiley-IEEE Press, 2009.

[29] C. M. Osburn and D. W. Ormond, "Sodium-induced barrier-height lowering and dielectric breakdown on SiO_2 films on silicon," *Journal of the Electrochemical Society*, vol. 121, no. 9, pp. 1195–1198, 1974.

[30] E. H. Snow, A. S. Grove, B. E. Deal *et al.*, "Ion transport phenomena in insulating films," *Journal of Applied Physics*, vol. 36, no. 5, pp. 1664–1673, 1965.

[31] T. H. Ning, C. M. Osburn, and H. N. Yu, "Electron trapping at positively charged centres in SiO_2," *Applied Physics Letters*, vol. 26, no. 5, pp. 248–250, 1975.

[32] T. H. DiStefano and M. Shatzkes, "Impact ionization model for dielectric instability and breakdown," *Applied Physics Letters*, vol. 25, no. 12, pp. 685–687, 1974.

[33] ——, "Dielectric instability and breakdown in SiO_2 thin films," *Journal of Vacuum Science and Technology*, vol. 13, no. 1, pp. 50–54, 1976.

[34] E. Harari, "Dielectric breakdown in electrically stressed thin films of thermal SiO_2 [MOS technology]," *Journal of Applied Physics*, vol. 49, no. 4, pp. 2478–2489, 1978.

[35] S. P. Li, "Measurement technique of time-dependant dielectric breakdown in MOS capacitors," *Microelectronics Reliability*, vol. 13, no. 3, pp. 209–214, 1974.

[36] E. S. Anolick, "Area and electrode effects on time dependent breakdowns through Si_3N_4+SiO_2 double layers," in *Proc. Electrochemical Society Spring Meeting*, 1975, pp. 131–132.

[37] D. L. Crook, "Method of determining reliability screens for Time Dependent Dielectric Breakdown," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1979, pp. 1–7.

[38] A. Berman, "Time-Zero dielectric reliability test by a ramp method," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1981, pp. 204–209.

[39] E. S. Anolick and G. R. Nelson, "Low field time dependent dielectric integrity," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1979, pp. 8–12.

[40] W. R. Hunter, "A failure rate based methodology for determining the maximum operating gate electric field, comprehending defect density and burn-in," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1996, pp. 37–43.

[41] ——, "The analysis of oxide reliability data," in *Final Report Integrated Reliability Workshop (IRW)*, 1998, pp. 114–134.

[42] J. W. McPherson and D. A. Baglee, "Acceleration factors for thin gate oxide stressing," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1985, pp. 1–5.

[43] J. W. McPherson, "Determination of the nature of molecular bonding in silica from time-dependent dielectric breakdown data," *Journal of Applied Physics*, vol. 95, no. 12, pp. 8101–8109, 2004.

[44] J. W. McPherson and R. B. Khamankar, "Molecular model for intrinsic time-dependent dielectric breakdown in SiO_2 dielectrics and the reliability implications for hyper-thin gate oxide," *Semiconductor Science and Technology*, vol. 15, no. 5, pp. 462–470, 2000.

[45] J. W. McPherson and H. C. Mogul, "Underlying physics of the thermochemical E model in describing low-field time-dependent dielectric breakdown in SiO_2 thin films," *Journal of Applied Physics*, vol. 84, no. 3, pp. 1513–1523, 1998.

[46] E. T. Ogawa, K. Jinyoung, G. S. Haase *et al.*, "Leakage, breakdown, and TDDB characteristics of porous low-k silica-based interconnect dielectrics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2003, pp. 166–172.

[47] F. Chen and M. Shinosky, "Addressing Cu/Low-k dielectric TDDB-reliability challenges for advanced CMOS technologies," *IEEE Transactions on Electron Devices*, vol. 56, no. 1, pp. 2–12, 2009.

[48] F. Chen, K. Chanda, J. Gill *et al.*, "Investigation of CVD SiCOH low-k time-dependent dielectric breakdown at 65nm node technology," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2005, pp. 501–507.

[49] N. Suzumura, S. Yamamoto, D. Kodama *et al.*, "A new TDDB degradation model based on Cu ion drift in Cu interconnect dielectrics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2006, pp. 484–489.

[50] F. Chen, O. Bravo, K. Chanda *et al.*, "A comprehensive study of low-k SiCOH TDDB phenomena and its reliability lifetime model development," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2006, pp. 46–53.

[51] I. C. Chen, S. Holland, and C. Hu, "A quantitative physical model for time-dependent breakdown in SiO_2," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1985, pp. 24–31.

[52] E. Wu, J. Sune, W. Lai *et al.*, "A comprehensive investigation of gate oxide breakdown of p+poly/pfets under inversion mode," in *Technical Digest, Int. Electron Devices Meeting (IEDM)*, 2005, pp. 396–399.

[53] J. R. Lloyd, E. Liniger, and T. M. Shaw, "Simple model for time-dependent dielectric breakdown in inter-and intralevel low-k dielectrics," *Journal of Applied Physics*, vol. 98, pp. 1–6, 2005.

[54] J. Su, I. Placencia, N. Barniol *et al.*, "On the breakdown statistics of very thin SiO_2 films," *Thin Solid Films*, vol. 185, no. 2, pp. 347–362, 1990.

[55] R. Degraeve, G. Groeseneken, R. Bellens *et al.*, "A consistent model for the thickness dependence of intrinsic breakdown in ultra-thin oxides," in *Technical Digest, Int. Electron Devices Meeting (IEDM)*, 1995, pp. 863 –866.

[56] J. H. Stathis, "Quantitative model of the thickness dependence of breakdown in ultra-thin oxides," *Microelectronic Engineering*, vol. 36, no. 1-4, pp. 325–328, 1997.

[57] C. Hong, L. Milor, and M. Z. Lin, "Analysis of the layout impact on electric fields in interconnect structures using finite element method," *Microelectronics Reliability*, vol. 44, no. 9-11, pp. 1867–1871, 2004.

[58] C. Hong and L. Milor, "Porosity-induced electric field enhancement and its impact on charge transport in porous inter-metal dielectrics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2006, pp. 679–680.

[59] C. A. Yuan, O. van der Sluis, G. Q. Zhang *et al.*, "Molecular simulation on the material/interfacial strength of the low-dielectric materials," *Microelectronics Reliability*, vol. 47, no. 9-11, pp. 1483–1491, 2007.

[60] Z. Chen, K. Prasad, C. Y. Li *et al.*, "Highly reliable dielectric/metal bilayer sidewall diffusion barrier in Cu/porous organic ultra low-k interconnects," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2004, pp. 320–325.

[61] S. H. Kim, K. T. Nam, A. Datta *et al.*, "Failure mechanism of a multilayer (TiN/Al/TiN) diffusion barrier between copper and silicon," *Journal of Applied Physics*, vol. 92, no. 9, pp. 5512–5519, 2002.

[62] J. Kim, E. T. Ogawa, and J. W. McPherson, "A statistical evaluation of the field acceleration parameter observed during time dependent dielectric breakdown testing of silica-based low-k interconnect dielectrics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2006, pp. 478–483.

[63] L. Milor and C. Hong, "Backend dielectric breakdown dependence on linewidth and pattern density," *Microelectronics Reliability*, vol. 47, no. 9-11, pp. 1473–1477, 2007.

[64] M. M. Bashir and L. Milor, "Analysis of the impact of linewidth variation on low-k dielectric breakdown," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2010, pp. 895–902.

[65] M. Vilmay, D. Roy, C. Monget *et al.*, "Copper-line topology impact on the reliability of SiOCH Low-k for the 45-nm technology node and beyond," *IEEE Transactions on Device and Materials Reliability*, vol. 9, no. 2, pp. 120–127, 2009.

[66] J. R. Lloyd, X. H. Liu, G. Bonilla *et al.*, "On the contribution of line-edge roughness to intralevel TDDB lifetime in low-k dielectrics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2009, pp. 602–605.

[67] J. W. Kim, N. H. Lee, H.-W. Kim *et al.*, "Intra-metal leakage reliability characteristics for line/via in copper/low-k interconnect structures," in *Final Report Integrated Reliability Workshop (IRW)*, 2002, pp. 36–40.

[68] J. Sung-Yup, K. Byoung-Joon, L. Nam Yeal *et al.*, "The characteristics of cu-drift induced dielectric breakdown under alternating polarity bias temperature stress," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2009, pp. 825–827.

[69] W. Liu, Y. Lim, F. Zhang *et al.*, "Effect of chemical mechanical polishing scratch on TDDB reliability and its reduction in 45nm BEOL process," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2009, pp. 613 –618.

[70] S. R. Nassif, A. J. Strojwas, and S. W. Director, "A methodology for worst-case analysis of integrated circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 5, no. 1, pp. 104–113, 1986.

[71] S. R. Nassif, "Modeling and forecasting of manufacturing variations," in *Proc. Asia and South Pacific Design Automation Conference (ASPDAC)*, 2001, pp. 145–149.

[72] D. K. Schroder, *Semiconductor material and device characterization*, 2nd ed. New York: Wiley, 1998.

[73] H. Wong, M. H. White, T. J. Krutsick *et al.*, "Modeling of transconductance degradation and extraction of threshold voltage in this oxide MOSFET's," *Solid-State Electronics*, vol. 30, no. 9, pp. 953–968, 1987.

[74] H. G. Lee, S. Y. Oh, and G. Fuller, "A simple and accurate method to measure the threshold voltage of an enhancement-mode MOSFET," *IEEE Transactions on Electron Devices*, vol. 29, no. 2, pp. 346–348, 1982.

[75] S. Borkar, T. Karnik, and V. K. De, "Design and reliability challenges in nanometer technologies," in *Proc. Design Automation Conference (DAC)*, 2004, p. 75.

[76] X. Tang, V. K. De, and J. D. Meindl, "Intrinsic MOSFET parameter fluctuations due to random dopant placement," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 5, no. 4, pp. 369–376, 1997.

[77] D. Burnett, K. Erington, C. Subramanian *et al.*, "Implications of fundamental threshold voltage variations for high-density SRAM and logic circuits," in *Digest of Technical Papers Symp. on VLSI Technology*, 1994, pp. 15–16.

[78] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, "Characterisation and modeling of mismatch in MOS transistors for precision analog design," *IEEE Journal of Solid-State Circuits*, vol. 21, no. 6, pp. 1057–1066, 1986.

[79] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, 1989.

[80] J. G. J. Chern, P. Chang, R. F. Motta *et al.*, "A new method to determine MOSFET channel length," *IEEE Electron Device Letters*, vol. 1, no. 9, pp. 170–173, 1980.

[81] K. Terada and H. Muta, "A new method to determine effective MOSFET channel length," *Japanese Journal of Applied Physics*, vol. 18, no. 5, pp. 953–959, 1979.

[82] Y. S. Jean and C. Y. Wu, "New extraction algorithm for the metallurgical channel length of conventional and LDD MOSFET's," *IEEE Transactions on Electron Devices*, vol. 43, no. 6, pp. 946–953, 1996.

[83] Y. Taur, D. S. Zicherman, D. R. Lombardi *et al.*, "A new 'shift and ratio' method for MOSFET channel-length extraction," *IEEE Electron Device Letters*, vol. 13, no. 5, pp. 267–269, 1992.

[84] F. J. Garcia Sanchez, A. Ortiz-Conde, M. Garcia Nunez *et al.*, "Extracting the series resistance and effective channel length of short-channel MOSFETs at liquid nitrogen temperature," *Solid-State Electronics*, vol. 37, no. 12, pp. 1943–1948, 1994.

[85] W. Fikry, G. Ghibaudo, H. Haddara *et al.*, "Method for extracting deep submicrometre MOSFET parameters," *Electronics Letters*, vol. 31, no. 9, pp. 762–764, 1995.

[86] C. C. McAndrew and P. A. Layman, "Mosfet effective channel length, threshold voltage, and series resistance determination by robust optimization," *IEEE Transactions on Electron Devices*, vol. 39, no. 10, pp. 2298–2311, 1992.

[87] P. R. Karlsson and K. O. Jeppson, "Efficient method for determining threshold voltage, series resistance and effective geometry of MOS transistors," *IEEE Transactions on Semiconductor Manufacturing*, vol. 9, no. 2, pp. 215–222, 1996.

[88] S. W. Lee, "A capacitance-based method for experimental determination of metallurgical channel length of submicron LDD MOSFETs," *IEEE Transactions on Electron Devices*, vol. 41, no. 3, pp. 403–412, 1994.

[89] B. J. Sheu and P. K. Ko, "Capacitance method to determine channel lengths for conventional and LDD MOSFET's," *IEEE Electron Device Letters*, vol. 5, no. 11, pp. 491–493, 1984.

[90] P. Vitanov, T. Dimitrova, R. Kamburova *et al.*, "Capacitance method for determination of LDD MOSFET geometrical parameters," *Solid-State Electronics*, vol. 35, no. 7, pp. 985–991, 1992.

[91] A. Borna, C. Progler, and D. Blaauw, "Correlation analysis of CD-variation and circuit performance under multiple sources of variability," in *Proc. of the SPIE*, vol. 5756, 2005, pp. 168–177.

[92] D. Sylvester and K. Keutzer, "Getting to the bottom of deep submicron," in *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, 1998, pp. 203–211.

[93] *International Technology Roadmap for Semiconductors*. Semiconductor Industry Association, 2007.

[94] D. S. Boning, K. Balakrishnan, C. Hong *et al.*, "Variation," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 1, pp. 63–71, 2008.

[95] S. A. Hareland, S. Krishnamurthy, S. Jallepalli *et al.*, "A computationally efficient model for inversion layer quantization effects in deep submicron N-channel MOSFETs," *IEEE Transactions on Electron Devices*, vol. 43, no. 1, pp. 90–96, 1996.

[96] H. C. De Graaff and F. M. Klaassen, *Compact transistor modelling for circuit design*. New York: Springer Verlag/Wein, 1990.

[97] G. Gildenblat, L. Xin, W. Wu *et al.*, "Psp: An advanced surface-potential-based MOSFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1979–1993, 2006.

[98] M. Miura-Mattausch, N. Sadachika, D. Navarro *et al.*, "HiSIM2: Advanced MOSFET Model Valid for RF Circuit Simulation," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1994–2007, 2006.

[99] J. R. Hauser, "A new and improved physics-based model for MOS transistors," *IEEE Transactions on Electron Devices*, vol. 52, no. 12, pp. 2640–2647, 2005.

[100] H. Katto, "A compact and accurate MOSFET model with simple expressions for linear, saturation and sub-threshold regions," *Solid-State Electronics*, vol. 50, no. 3, pp. 301–308, 2006.

[101] C. C. McAndrew, "Statistical modeling for circuit simulation," in *Proc. Int. Symp. on Quality Elec. Design (ISQED)*, 2003, pp. 357–362.

[102] S. R. Nassif, "Design for variability in DSM technologies [deep submicron technologies]," in *Proc. Int. Symp. on Quality Elec. Design (ISQED)*, 2000, pp. 451–454.

[103] S. G. Narendra and A. P. Chandrakasan, *Leakage in nanometer CMOS technologies*, ser. Series on integrated circuits and systems. New York: Springer, 2006.

[104] A. Agarwal, D. Blaauw, and V. Zolotov, "Statistical timing analysis for intra-die process variations with spatial correlations," in *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, 2003, pp. 900–907.

[105] S. Nassif, "Delay variability: sources, impacts and trends," in *Digest of Technical Papers Int. Solid-State Circuits Conf. (ISSCC)*, 2000, pp. 368–369.

[106] N. Drego, A. Chandrakasan, and D. Boning, "Lack of spatial correlation in mosfet threshold voltage variation and implications for voltage scaling," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 2, pp. 245–255, 2009.

[107] ——, "A test-structure to efficiently study threshold-voltage variation in large MOSFET arrays," in *Proc. Int. Symp. on Quality Elec. Design (ISQED)*, 2007, pp. 281–286.

[108] W. Zhao, Y. Cao, F. Liu *et al.*, "Rigorous extraction of process variations for 65nm CMOS design," in *Proc. European Solid-State Device Research Conf.*, 2007, pp. 89–92.

[109] K. Agarwal and S. Nassif, "The impact of random device variation on SRAM cell stability in sub-90-nm cmos technologies," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 16, no. 1, pp. 86–97, 2008.

[110] K. Nowka, S. Nassif, and K. Agarwal, "Characterization and design for variability and reliability," in *Proc. Custom Integrated Circuits Conf. (CICC)*, 2008, pp. 341–346.

[111] K. Gettings and D. S. Boning, "Study of CMOS process variation by multiplexing analog characteristics," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 4, pp. 513–525, 2008.

[112] V. Wang, K. Agarwal, S. R. Nassif *et al.*, "A simplified design model for random process variability," *IEEE Transactions on Semiconductor Manufacturing*, vol. 22, no. 1, pp. 12–21, 2009.

[113] F. Chen, F. Ungar, A. H. Fischer *et al.*, "Reliability characterization of BEOL vertical natural capacitor using copper and low-k sicoh dielectric for 65nm RF and mixed-signal applications," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2006, pp. 490–495.

[114] F. Chen, P. McLaughlin, J. Gambino *et al.*, "The effect of metal area and line spacing on TDDB characteristics of 45nm Low-k SiCOH dielectrics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2007, pp. 382–389.

[115] D. R. Wolters and J. F. Verwey, "Breakdown and wear-out phenomena in SiO_2 films," in *Instabilities in Silicon Devices*, G. Barbottin and A. Vapaille, Eds. Amsterdam, The Netherlands: North-Holland, 1986, pp. 315–362.

[116] E. Y. Wu, W. W. Abadeer, L. K. Han *et al.*, "Challenges for accurate reliability projections in the ultra-thin oxide regime," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 1999, pp. 57–65.

[117] E. Y. Wu and R. P. Vollertsen, "On the weibull shape factor of intrinsic breakdown of dielectric films and its accurate experimental determination. Part I: theory, methodology, experimental techniques," *IEEE Transactions on Electron Devices*, vol. 49, no. 12, pp. 2131–2140, 2002.

[118] G. S. Haase, E. T. Ogawa, and J. W. McPherson, "Reliability analysis method for low-k interconnect dielectrics breakdown in integrated circuits," *Journal of Applied Physics*, vol. 98, no. 3, p. 034503, 2005.

[119] S.-C. Lee, A. S. Oates, and K. M. Chang, "Limitation of low-k reliability due to dielectric breakdown at vias," in *Proc. IEEE Int. Iterconnect Technology Conference (IITC)*, 2008, pp. 177–179.

[120] C. J. Mogab, "The loading effect in plasma etching," *Journal of the Electrochemical Society*, vol. 124, no. 8, pp. 1262–1268, 1977.

[121] A. Misaka, K. Harafuji, H. Nakagawa *et al.*, "A simulation of micro-loading phenomena in dry-etching process using a new adsorption model," in *Technical Digest, Int. Electron Devices Meeting (IEDM)*. IEEE, 1993, pp. 857–860.

[122] M. Bashir and L. Milor, "A methodology to extract failure rates for low-k dielectric breakdown with multiple geometries and in the presence of die-to-die linewidth variation," *Microelectronics Reliability*, vol. 49, no. 9-11, pp. 1096–1102, 2009.

[123] ——, "Modeling low-k dielectric breakdown to determine lifetime requirements," *IEEE Design & Test of Computers*, vol. 26, no. 6, pp. 18–27, 2009.

[124] T. M. Jeong, S. W. Choi, J. R. Park *et al.*, "Flare in microlithographic exposure tools," *Japanese Journal of Applied Physics, Part 1:Regular Papers, Short Notes & Review Papers*, vol. 1, no. 41, pp. 5113–5119, 2002.

[125] K. O. Abrokwah, P. R. Chidambaram, and D. S. Boning, "Pattern based prediction for plasma etch," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 2, pp. 77–86, 2007.

[126] H. K. Taylor, S. Hongwei, T. F. Hill *et al.*, "Characterizing and predicting spatial nonuniformity in the deep reactive ion etching of silicon," *Journal of the Electrochemical Society*, vol. 153, no. 8, pp. 575–585, 2006.

[127] S. Jensen, O. Hansen, and M. C. Mic, "Inverse microloading effect in reactive ion etching of silicon," in *Proc. of the Electrochemical Society International Symposium on Plasma Processing XIV*, 2002, pp. 218–226.

[128] V. Bliznetsov, M. L. Chua, M. M. Roy *et al.*, "Challenges of pattern transfer for ultra-low-k OSG film Aurora(TM)ULK," *Thin Solid Films*, vol. 462-463, pp. 235–239, 2004.

[129] R. A. Gottscho, C. W. Jurgensen, and D. J. Vitkavage, "Microscopic uniformity in plasma etching," *Journal of Vacuum Science and Technology*, vol. 10, no. 5, pp. 2133–2147, 1992.

[130] G. Whl, M. Matthes, and A. Weisheit, "Reactive ion etching of deep trenches in silicon with CF_2Cl_2 and O_2," *Vacuum*, vol. 38, no. 11, pp. 1011–1014, 1988.

[131] C. Ye, Y. Xu, X. Huang *et al.*, "Effect of low-frequency power on etching of SiCOH low-k films in CHF_3 $13.56MHz/2MHz$ dual-frequency capacitively coupled plasma," *Microelectronic Engineering*, vol. 86, no. 3, pp. 421–424, 2009.

[132] F. Chen, O. Bravo, D. Harmon *et al.*, "Cu/low-k dielectric TDDB reliability issues for advanced CMOS technologies," *Microelectronics Reliability*, vol. 48, no. 8-9, pp. 1375–1383, 2008.

[133] C. U. Kim, J. Y. Park, N. L. Michael *et al.*, "Detection of barrier failure in interconnects with ultralow-k materials: ICV voltammogram," in *Proc. Advanced Metallization Conference*, 2005, pp. 679–685.

[134] C. J. Zhai, H. W. Yao, A. P. Marathe *et al.*, "Simulation and experiments of stress migration for Cu/low-k BEoL," *IEEE Transactions on Device and Materials Reliability*, vol. 4, no. 3, pp. 523–529, 2004.

[135] R. Kumar, T. K. S. Wong, B. R. Murthy *et al.*, "Effects of plasma treatments on ultralow-k dielectric film and Ta barrier properties in Cu damascene processing," *Journal of the Electrochemical Society*, vol. 153, no. 5, pp. 420–427, 2006.

[136] E. Vinogradova, E. Osei-Yiadom, C. E. Smith *et al.*, "Effects of plasmas on porous low dielectric constant CVD SiOCH films," *Microelectronic Engineering*, vol. 86, no. 2, pp. 176–180, 2009.

[137] M. Juppo, P. Aln, M. Ritala *et al.*, "Atomic layer deposition of titanium nitride thin films using tert-butylamine and allylamine as reductive nitrogen sources," *Electrochemical and Solid-State Letters*, vol. 5, pp. C4–C6, 2002.

[138] C. Ryu, H. Lee, K.-W. Kwon *et al.*, "Barriers for copper interconnections," *Solid state technology*, vol. 42, no. 4, pp. 53–54, 1999.

[139] F. Chen, J. R. Lloyd, K. Chanda *et al.*, "Line edge roughness and spacing effect on low-k TDDB characteristics," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2008, pp. 132–137.

[140] M. Vilmay, D. Roy, C. Monget *et al.*, "Copper line topology impact on the SiOCH low-k reliability in sub $45nm$ technology node. from the time-dependent dielectric breakdown to the product lifetime," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2009, pp. 606–612.

[141] T. Pompl, C. Schlunder, M. Hommel *et al.*, "Practical aspects of reliability analysis for ic designs," in *Proc. Design Automation Conference (DAC)*, 2006, pp. 193–198.

[142] G. S. Haase and J. W. McPherson, "Modeling of interconnect dielectric lifetime under stress conditions and new extrapolation methodologies for time-dependent dielectric breakdown," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2007, pp. 390–398.

[143] J. Kim, E. T. Ogawa, and J. W. McPherson, "Time dependent dielectric breakdown characteristics of low-k dielectric (SiOC) over a wide range of test areas and electric fields," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2007, pp. 399–404.

[144] Synopsys, Inc., "FAMMOS," http://www.synopsys.com.

[145] S. Alam, C. Gan, C. Thompson *et al.*, "Reliability computer-aided design tool for full-chip electromigration analysis and comparison with different interconnect metallizations," *Microelectronics Journal*, vol. 38, no. 4-5, pp. 463 – 473, 2007.

[146] C. Hu, "IC reliability simulation," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 3, pp. 241–246, 1992.

[147] X. Li, J. Qin, B. Huang *et al.*, "A new SPICE reliability simulation method for deep submicrometer CMOS VLSI circuits," *IEEE Transactions on Device and Materials Reliability*, vol. 6, no. 2, pp. 247–257, 2006.

[148] M. Bashir, L. Milor, D. H. Kim *et al.*, "Methodology to determine the impact of linewidth variation on chip scale copper/low-k backend dielectric breakdown," *Microelectronics Reliability*, vol. 50, no. 9-11, pp. 1341–1346, 2010.

[149] M. Bashir and L. Milor, "Towards a chip level reliability simulator for copper/low-k backend processes," in *Proc. Design, Automation & Test in Europe (DATE)*, 2010, pp. 279–282.

[150] M. Bashir, D. H. Kim, S. K. Lim *et al.*, "Tddb chip reliability in copper interconnects," in *Final Report Integrated Reliability Workshop (IRW)*, 2010.

[151] NCSU EDA, "NCSU FreePDK45," http://www.eda.ncsu.edu/wiki/FreePDK.

[152] Launchbird Design Systems, Inc., "CF FFT," http://www.opencores.org.

[153] Synopsys, Inc., "Design Compiler," http://www.synopsys.com.

[154] Cadence Design Systems, Inc., "Soc Encounter RTL-to-GDSII System 8.1," http://www.cadence.com.

[155] Synopsys, Inc., "PrimeTime," http://www.synopsys.com.

[156] M. Bashir, D. H. Kim, K. Athikulwongse *et al.*, "Backend Low-k TDDB chip reliability simulator," in *Proc. Int. Reliability Physics Symp. (IRPS)*, 2011.

[157] J. E. Jackson, *A user's guide to principal components*. Wiley and Sons, 1978.

[158] C. Michael and M. Ismail, "Statistical modeling of device mismatch for analog MOS integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 2, pp. 154–166, 1992.

[159] M. Orshansky, L. Milor, C. Pinhong *et al.*, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, no. 5, pp. 544–553, 2002.

[160] J. Pineda de Gyvez and H. P. Tuinhout, "Threshold voltage mismatch and intra-die leakage current in digital cmos circuits," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 1, pp. 157–168, 2004.

[161] C. H. Ge, C. C. Lin, C. H. Ko *et al.*, "Process-strained Si (PSS) CMOS technology featuring 3D strain engineering," in *Technical Digest, Int. Electron Devices Meeting (IEDM)*, 2003, pp. 3.7.1–3.7.4.

[162] P. R. Chidambaram, C. Bowen, S. Chakravarthi *et al.*, "Fundamentals of silicon material properties for successful exploitation of strain engineering in modern CMOS manufacturing," *IEEE Transactions on Electron Devices*, vol. 53, no. 5, pp. 944–964, 2006.

[163] L. H. A. Leunissen, R. Jonckheere, K. Ronse *et al.*, "Influence of gate patterning on line edge roughness," *Journal of Vacuum Science & Technology B (Microelectronics and Nanometer Structures)*, vol. 21, no. Copyright 2003, IEE, pp. 3140–3143, 2003.

[164] A. Yabata, O. Koike, J. Hashimoto *et al.*, "New mechanism of LER formation in gate process," in *IEEE Int. Symp. on Semiconductor Manufacturing (ISSM)*, 2006, pp. 99–102.

[165] P. Friedberg, Y. Cao, J. Cain *et al.*, "Modeling within-field gate length spatial variation for process-design co-optimization," in *Proc. of the SPIE*, vol. 5756. SPIE, 2005, pp. 178–188.

[166] J. Xiong, V. Zolotov, and H. Lei, "Robust extraction of spatial correlation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 4, pp. 619–631, 2007.

[167] A. Singhee, S. Singhal, and R. A. Rutenbar, "Exploiting correlation kernels for efficient handling of intra-die spatial correlation, with application to statistical timing," in *Proc. Design, Automation & Test in Europe (DATE)*, 2008, pp. 856–861.

[168] Q. Fu, L. Wai-Shing, T. Jun *et al.*, "Characterizing intra-die spatial correlation using spectral density method," in *Proc. Int. Symp. on Quality Elec. Design (ISQED)*, 2008, pp. 718–723.

[169] H. J. Levinson, *Principles of lithography*, 2nd ed. SPIE Press, 2005.

[170] L. Chen, L. S. Milor, C. H. Ouyang *et al.*, "Analysis of the impact of proximity correction algorithms on circuit performance," *IEEE Transactions on Semiconductor Manufacturing*, vol. 12, no. 3, pp. 313–322, 1999.

[171] J. P. Cain, H. Zhang, and C. J. Spanos, "Optimum sampling for characterization of systematic variation in photolithography," in *Proc. Metrology, Inspection, and Process Control for Microlithography XVI*, vol. 4689. Santa Clara, CA, USA: SPIE, 2002, pp. 430–442.

[172] M. Orshansky, L. Milor, and C. Hu, "Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial mask-level correction," *IEEE Transactions on Semiconductor Manufacturing*, vol. 17, no. 1, pp. 2–11, 2004.

[173] C. A. Mack, "Measuring and modeling flare in optical lithography," in *Proc. Optical Microlithography XVI*, vol. 5040, 2003, pp. 151–161.

[174] M. Osawa, T. Yao, H. Aoyama *et al.*, "Correction for local flare effects approximated with double gaussian profile in arf lithography," *Journal of Vacuum Science & Technology B (Microelectronics and Nanometer Structures)*, vol. 21, pp. 2806–2809, 2003.

[175] S. C. Abraham, "A new chemistry for a high-density plasma etcher that improves etch rate loading on the TiN ARC layer when geometries are below 0.5 micron," in *Proc. Advanced Semiconductor Manufacturing Conference and Workshop*, 1996, pp. 328–332.

[176] B. E. Stine, D. O. Ouma, R. R. Divecha *et al.*, "Rapid characterization and modeling of pattern-dependent variation in chemical-mechanical polishing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 11, no. 1, pp. 129–140, 1998.

[177] C. Ouyang, K. Ryu, L. Milor *et al.*, "An analytical model of multiple ILD thickness variation induced by interaction of layout pattern and CMP process," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, no. 3, pp. 286–292, 2000.

[178] S. Lakshminarayanan, P. J. Wright, and J. Pallinti, "Electrical characterization of the copper CMP process and derivation of metal layout rules," *IEEE Transactions on Semiconductor Manufacturing*, vol. 16, no. 4, pp. 668–676, 2003.

[179] M. Choi and L. Milor, "Impact on circuit performance of deterministic within-die variation in nanoscale semiconductor manufacturing," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 7, pp. 1350–1367, 2006.

[180] D. Markovic, V. Stojanovic, B. Nikolic *et al.*, "Methods for true energy-performance optimization," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 8, pp. 1282–1293, 2004.

[181] B. Cline, K. Chopra, D. Blaauw *et al.*, "Analysis and modeling of CD variation for statistical static timing," in *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, 2006, pp. 60–66.

[182] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, 2002.

[183] K. Agarwal, J. Hayes, and S. Nassif, "Fast characterization of threshold voltage fluctuation in MOS devices," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 4, pp. 526–533, 2008.

[184] H. Chang and S. S. Sapatnekar, "Statistical timing analysis under spatial correlations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 9, pp. 1467–1482, 2005.

[185] D. Blaauw, K. Chopra, A. Srivastava *et al.*, "Statistical timing analysis: From basic principles to state of the art," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 4, pp. 589–607, 2008.

[186] S. Bhardwaj, S. Vrudhula, and A. Goel, "A unified approach for full chip statistical timing and leakage analysis of nanoscale circuits considering intradie process variations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 10, pp. 1812–1825, 2008.

[187] P. Gupta and F.-L. Heng, "Toward a systematic-variation aware timing methodology," in *Proc. Design Automation Conference (DAC)*, 2004, pp. 321–326.

[188] A. Agarwal, D. Blaauw, V. Zolotov *et al.*, "Statistical delay computation considering spatial correlations," in *Proc. Asia and South Pacific Design Automation Conference (ASPDAC)*, 2003, pp. 271–276.

[189] M. Yamamoto, H. Endo, and H. Masuda, "Development of a large-scale TEG for evaluation and analysis of yield and variation," *IEEE Transactions on Semiconductor Manufacturing*, vol. 17, no. 2, pp. 111–122, 2004.

[190] S. Ohkawa, M. Aoki, and H. Masuda, "Analysis and characterization of device variations in an LSI chip using an integrated device matrix array," *IEEE Transactions on Semiconductor Manufacturing*, vol. 17, no. 2, pp. 155–165, 2004.

[191] K. Agarwal, F. Liu, C. McDowell *et al.*, "A test structure for characterizing local device mismatches," in *IEEE Symp. on VLSI Circuits*, 2006, pp. 67–68.

[192] Y. Tsividis, *Operation and modeling of the MOS transistor*, 2nd ed. Boston: WCB/McGraw-Hill, 1999.

[193] Y. Taur and T. H. Ning, *Fundamentals of modern VLSI devices.* Cambrige, UK: Cambridge University Press, 1998.

[194] N. Arora, *Mosfet modeling for VLSI simulation : theory and practice.* New Jersey: World Scientific, 2007.

[195] A. Agarwal, V. Zolotov, and D. Blaauw, "Statistical timing analysis using bounds and selective enumeration," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 9, pp. 1243 – 1260, 2003.

[196] X. Li, J. Le, M. Celik *et al.*, "Defining statistical timing sensitivity for logic circuits with large-scale process and environmental variations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 6, pp. 1041 –1054, 2008.

[197] L. Zhang, C. Weijen, H. Yuhen *et al.*, "Correlation-preserved non-gaussian statistical timing analysis with quadratic timing model," in *Proc. Design Automation Conference (DAC)*, 2005, pp. 83–88.

[198] V. Khandelwal and A. Srivastava, "A general framework for accurate statistical timing analysis considering correlations," in *Proc. Design Automation Conference (DAC)*, 2005, pp. 89–94.

[199] J. Singh and S. S. Sapatnekar, "A scalable statistical static timing analyzer incorporating correlated non-gaussian and gaussian parameter variations," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 1, pp. 160–173, 2008.

VITA

Muhammad Muqarrab Bashir was born in Pakistan on July the 15^{th}, 1983. He received the BSc degree in electrical engineering from the University of Engineering and Technology (UET), Lahore, Pakistan, in 2006 and the M.S. degree in electrical and computer engineering from the Georgia Institute of Technology in 2008. His research interests include semiconductors, device physics, advanced materials for semiconductors, reliability of Copper interconnects, system level impact of Copper interconnect reliability, reliability statistics, device modeling, compact models and modeling the impact of variation on VLSI circuits.